(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 February 2004 (26.02.2004)

PCT

(10) International Publication Number
WO 2004/016760 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number:
PCT/US2003/025625

(22) International Filing Date: 15 August 2003 (15.08.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/404,063	16 August 2002 (16.08.2002)	US
60/456,368	21 March 2003 (21.03.2003)	US
60/458,853	27 March 2003 (27.03.2003)	US
60/458,696	27 March 2003 (27.03.2003)	US

(71) Applicant *(for all designated States except US)*: **GENENCOR INTERNATIONAL, INC.** [US/US]; 925 Page Mill Road, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **DAY, Anthony, G.** [GB/US]; 551 Monterey Boulevard, San Francisco, CA 94127 (US). **GOEDEGEBUUR, Frits** [NL/NL]; Rozenlaan 128, NL-3135XV Vlaardingen (NL). **GUALFETTI, Peter** [US/US]; 129 Arkansas Street #1, San Francisco, CA 94107 (US). **MITCHINSON, Colin** [GB/US]; 381 Myrtle Street, Half Moon Bay, CA 94019 (US). **NEEFE, Paulien** [NL/NL]; Basalt 48, NL-2719 TN Zoetermeer (NL). **SANDGREN, Mats** [SE/SE]; Rabeniusvagen 20, S-SE-75655 Uppsala (SE). **SHAW, Andrew** [GB/US]; 2560 Hyde Street, San Francisco, CA 94109 (US). **STAHLBERG, Jerry** [SE/SE]; Malma Skogsvag 3A, S-SE-75645 Uppsala (SE).

(74) Agent: **BOYD, Victoria, L.**; Genencor International, INC., 925 Page Mill Road, Palo Alto, CA 94304 (US).

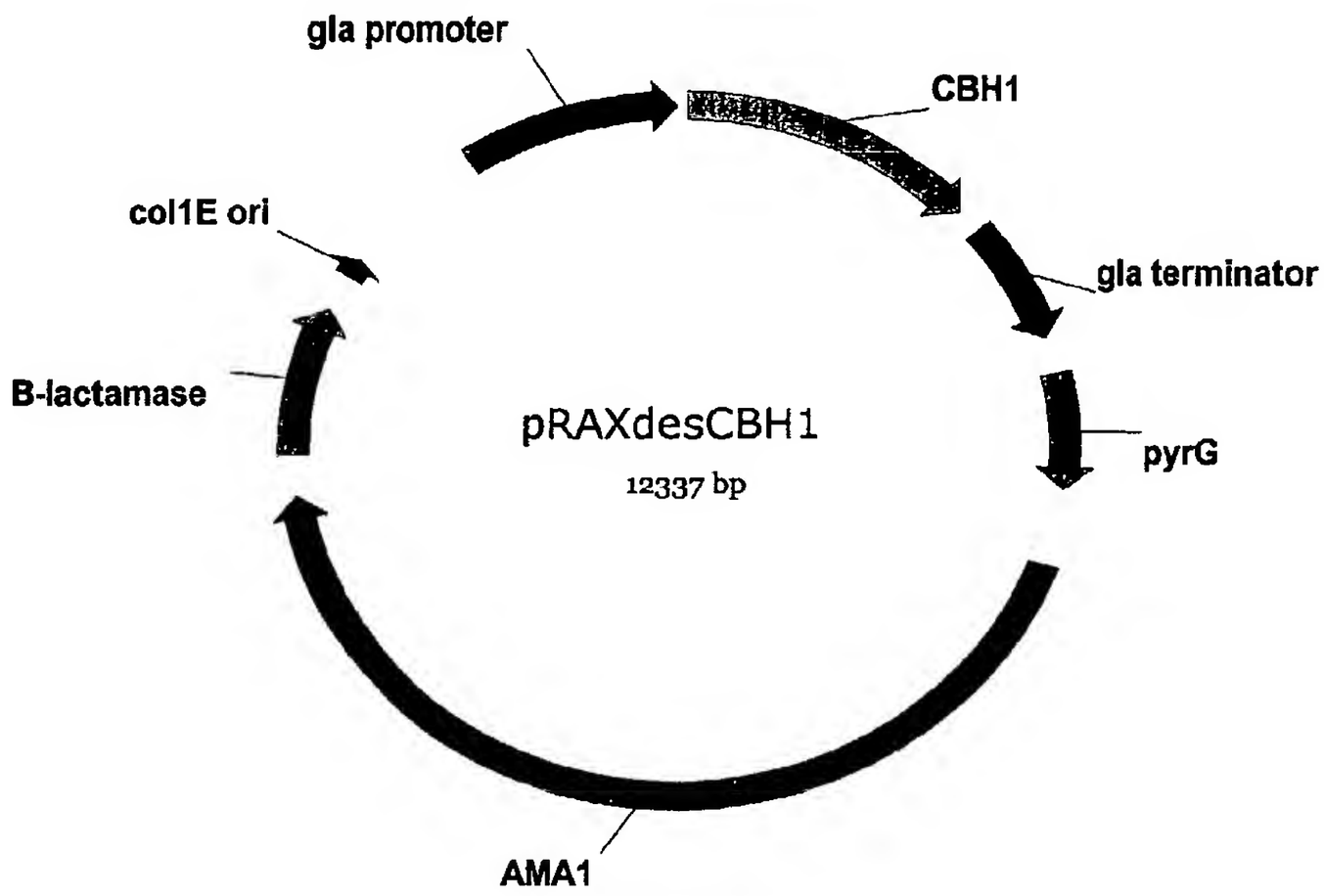
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

[Continued on next page]

(54) Title: NOVEL VARIANT *HYPROCREA JECORINA* CBH1 CELLULASES

Replicative expression pRAXdesCBH1 vector of CBH1 genes under the control of the glucoamylase promotor.



(57) Abstract: Described herein are variants of *H. jecorina* CBH I, a Cel7 enzyme. The present invention provides novel cellobiohydrolases that have improved thermostability and reversibility.

WO 2004/016760 A2

European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

NOVEL VARIANT

HYPROCREA JECORINA CBH1 CELLULASES

CROSS-REFERENCE TO RELATED APPLICATIONS

[01] This application claims priority to U.S. Provisional Application No. 60/404,063, filed August 16, 2002 (Attorney Docket No. GC772P), to U.S. Provisional Application No. 60/458,853 filed March 27, 2003 (Attorney Docket No. GC772-2P), to U.S. Provisional Application No. 60/456,368 filed March 21, 2003 (Attorney Docket No. GC793P) and to U.S. Provisional Application No. 60/458,696 filed March 27, 2003 (Attorney Docket No. GC793-2P), all herein incorporated by reference.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

[02] Portions of this work were funded by Subcontract No. ZCO-0-30017-01 with the National Renewable Energy Laboratory under Prime Contract No. DE-AC36-99GO10337 with the U.S. Department of Energy. Accordingly, the United States Government may have certain rights in this invention.

FIELD OF THE INVENTION

[03] The present invention relates to variant cellobiohydrolase enzymes and isolated nucleic acid sequences which encode polypeptides having cellobiohydrolase activity. The invention also relates to nucleic acid constructs, vectors, and host cells comprising the nucleic acid sequences as well as methods for producing recombinant variant CBH polypeptides.

REFERENCES

1. Sheehan and Himmel *Biotechnology Progress 15, pp* 817-827 (1999)
2. Matti Linko Proceedings of the Second TRICEL Symposium on *Trichoderma reesei* Cellulases and Other Hydrolases *pp* 9-11 (1993)
3. Tuula T. Teeri *Trends in Biotechnology 15, pp* 160-167 (1997)
4. T.T. Teeri et al. Spec. Publ. - R. Soc. Chem., 246 (Recent Advances in Carbohydrate Bioengineering), *pp* 302-308. (1999)
5. PDB reference **2OVW**: Sulzenbacher, G., Schulein, M., Davies, G. J. *Biochemistry* 36 *pp.* 5902 (1997)
PDB reference **1A39**: Davies, G. J., Ducros, V., Lewis, R. J., Borchert, T. V., Schulein, M. *Journal of Biotechnology* 57 *pp.* 91 (1997)
7. PDB reference **6CEL**: Divne, C., Stahlberg, J., Teeri, T. T., Jones, T. A. *Journal of Molecular Biology* 275 *pp.* 309 (1998)

8. PDB reference **1EG1**: Kleywegt, G. J., Zou, J. Y., Divne, C., Davies, G. J., Sinning, I., Stahlberg, J., Reinikainen, T., Srisodsuk, M., Teeri, T. T., Jones, T. A. *Journal of Molecular Biology* 272 *pp.* 383 (1997)

9. PDB reference **1DY4 (8CEL)**: J. Stahlberg, H. Henriksson, C. Divne, R. Isaksson, G. Pettersson, G. Johansson, T. A. Jones

BACKGROUND OF THE INVENTION

[04] Cellulose and hemicellulose are the most abundant plant materials produced by photosynthesis. They can be degraded and used as an energy source by numerous microorganisms, including bacteria, yeast and fungi, that produce extracellular enzymes capable of hydrolysis of the polymeric substrates to monomeric sugars (Aro *et al.*, J. Biol. Chem., vol. 276, no. 26, pp. 24309-24314, June 29, 2001). As the limits of non-renewable resources approach, the potential of cellulose to become a major renewable energy resource is enormous (Krishna *et al.*, Bioresource Tech. 77:193-196, 2001). The effective utilization of cellulose through biological processes is one approach to overcoming the shortage of foods, feeds, and fuels (Ohmiya *et al.*, Biotechnol. Gen. Engineer. Rev. vol. 14, pp. 365-414, 1997).

[05] Cellulases are enzymes that hydrolyze cellulose (beta-1,4-glucan or beta D-glucosidic linkages) resulting in the formation of glucose, cellobiose, cellooligosaccharides, and the like. Cellulases have been traditionally divided into three major classes: endoglucanases (EC 3.2.1.4) ("EG"), exoglucanases or cellobiohydrolases (EC 3.2.1.91) ("CBH") and beta-glucosidases ([beta] -D-glucoside glucohydrolase; EC 3.2.1.21) ("BG"). (Knowles *et al.*, TIBTECH 5, 255-261, 1987; Shulein, Methods Enzymol., 160, 25, pp. 234-243, 1988). Endoglucanases act mainly on the amorphous parts of the cellulose fibre, whereas cellobiohydrolases are also able to degrade crystalline cellulose (Nevalainen and Penttila, Mycota, 303-319, 1995). Thus, the presence of a cellobiohydrolase in a cellulase system is required for efficient solubilization of crystalline cellulose (Suurnakki, *et al.* Cellulose 7:189-209, 2000). Beta-glucosidase acts to liberate D-glucose units from cellobiose, cello-oligosaccharides, and other glucosides (Freer, J. Biol. Chem. vol. 268, no. 13, pp. 9337-9342, 1993).

[06] Cellulases are known to be produced by a large number of bacteria, yeast and fungi. Certain fungi produce a complete cellulase system capable of degrading crystalline forms of cellulose, such that the cellulases are readily produced in large quantities via fermentation. Filamentous fungi play a special role since many yeast, such as *Saccharomyces cerevisiae*, lack the ability to hydrolyze cellulose. See, *e.g.*, Aro *et al.*,

2001; Aubert *et al.*, 1988; Wood *et al.*, Methods in Enzymology, vol. 160, no. 9, pp. 87-116, 1988, and Coughlan, *et al.*, "Comparative Biochemistry of Fungal and Bacterial Cellulolytic Enzyme Systems" Biochemistry and Genetics of Cellulose Degradation, pp. 11-30 1988..

[07] The fungal cellulase classifications of CBH, EG and BG can be further expanded to include multiple components within each classification. For example, multiple CBHs, EGs and BGs have been isolated from a variety of fungal sources including *Trichoderma* reesei which contains known genes for 2 CBHs, *i.e.*, CBH I and CBH II, at least 8 EGs, *i.e.*, EG I, EG II , EG III, EGIV, EGV, EGVI, EGVII and EGVIII, and at least 5 BGs, i.e., BG1, BG2, BG3, BG4 and BG5.

[08] In order to efficiently convert crystalline cellulose to glucose the complete cellulase system comprising components from each of the CBH, EG and BG classifications is required, with isolated components less effective in hydrolyzing crystalline cellulose (Filho *et al.*, Can. J. Microbiol. 42:1-5, 1996). A synergistic relationship has been observed between cellulase components from different classifications. In particular, the EG-type cellulases and CBH- type cellulases synergistically interact to more efficiently degrade cellulose. See, *e.g.*, Wood, Biochemical Society Transactions, 611th Meeting, Galway, vol. 13, pp. 407-410, 1985.

[09] Cellulases are known in the art to be useful in the treatment of textiles for the purposes of enhancing the cleaning ability of detergent compositions, for use as a softening agent, for improving the feel and appearance of cotton fabrics, and the like (Kumar *et al.*, Textile Chemist and Colorist, 29:37-42, 1997).

[10] Cellulase-containing detergent compositions with improved cleaning performance (US Pat. No. 4,435,307; GB App. Nos. 2,095,275 and 2,094,826) and for use in the treatment of fabric to improve the feel and appearance of the textile (US Pat. Nos. 5,648,263, 5,691,178, and 5,776,757; GB App. No. 1,358,599; The Shizuoka Prefectural Hammamatsu Textile Industrial Research Institute Report, Vol. 24, pp. 54-61, 1986), have been described.

[11] Hence, cellulases produced in fungi and bacteria have received significant attention. In particular, fermentation of *Trichoderma spp.* (*e.g.*, *Trichoderma longibrachiatum* or *Trichoderma reesei*) has been shown to produce a complete cellulase system capable of degrading crystalline forms of cellulose.

[12] Although cellulase compositions have been previously described, there remains a need for new and improved cellulase compositions for use in household detergents,

stonewashing compositions or laundry detergents, etc. Cellulases that exhibit improved performance are of particular interest.

BRIEF SUMMARY OF THE INVENTION

[13] The invention provides an isolated cellulase protein, identified herein as variant CBH I, and nucleic acids which encode a variant CBH I.

[14] In one embodiment the invention is directed to a variant CBH I cellulase, wherein said variant comprises a substitution or deletion at a position corresponding to one or more of residues S8, Q17, G22, T41, N49, S57, N64, A68, A77, N89, S92, N103, A112, S113, E193, S196, M213, L225, T226, P227, T246, D249, R251, Y252, T255, D257, D259, S278, S279, K286, L288, E295, T296, S297, A299, N301, E325, T332, F338, S342, F352, T356, Y371, T380, Y381, V393, R394, S398, V403, S411, G430, G440, T445, T462, T484, Q487, and P491 in CBH I from *Hypocrea jecorina* (SEQ ID NO: 2). In first aspect, the invention encompasses an isolated nucleic acid encoding a polypeptide having cellobiohydrolase activity, which polypeptide is a variant of a glycosyl hydrolase of family 7, and wherein said nucleic acid encodes a substitution at a residue which is sensitive to temperature stress in the polypeptide encoded by said nucleic acid, wherein said variant cellobiohydrolase is derived from *H. jecorina* cellobiohydrolase. In second aspect, the invention encompasses an isolated nucleic acid encoding a polypeptide having cellobiohydrolase activity, which polypeptide is a variant of a glycosyl hydrolase of family 7, and wherein said nucleic acid encodes a substitution at a residue which is effects enzyme processitivity in the polypeptide encoded by said nucleic acid, wherein said variant cellobiohydrolase is derived from *H. jecorina* cellobiohydrolase. In third aspect, the invention encompasses an isolated nucleic acid encoding a polypeptide having cellobiohydrolase activity, which polypeptide is a variant of a glycosyl hydrolase of family 7, and wherein said nucleic acid encodes a substitution at a residue which is effects product inhibition in the polypeptide encoded by said nucleic acid, wherein said variant cellobiohydrolase is derived from *H. jecorina* cellobiohydrolase.

[15] In a second embodiment the invention is directed to a variant CBH I cellulose comprising a substitution at a position corresponding to one or more of residues S8P, Q17L, G22D, T41I, N49S, S57N, N64D, A68T, A77D, N89D, S92T, N103I, A112E, S113(T/N/D), E193V, S196T, M213I, L225F, T226A, P227(L/T/A), T246(C/A), D249K, R251A, Y252(A/Q), T255P, D257E, D259W, S278P, S279N, K286M, L288F, E295K, T296P, S297T, A299E, N301(R/K), E325K, T332(K/Y/H), F338Y, S342Y, F352L, T356L, Y371C, T380G, Y381D, V393G, R394A, S398T, V403D, S411F, G430F, G440R, T462I,

T484S, Q487L and/or P491L in CBH I from *Hypocrea jecorina* (SEQ ID NO: 2). In one aspect of this embodiment the variant CBH I cellulase further comprises a deletion at a position corresponding to T445 in CBH I from *Hypocrea jecorina* (SEQ ID NO: 2). In a second aspect of this embodiment the variant CBH I cellulase further comprises the deletion of residues corresponding to residues 382-393 in CBH I of *Hypocrea jecorina* (SEQ ID NO: 2).

[16] In a third embodiment the invention is directed to a variant CBH I cellulase, wherein said variant comprises a substitution at a position corresponding to a residue selected from the group consisting of S8P, N49S, A68T, A77D, N89D, S92T, S113(N/D), L225F, P227(A/L/T), D249K, T255P, D257E, S279N, L288F, E295K, S297T, A299E, N301K, T332(K/Y/H), F338Y, T356L, V393G, G430F in CBH I from *Hypocrea jecorina* (SEQ ID NO: 2).

[17] In a fourth embodiment the invention is directed to a variant CBH I consists essentially of the mutations selected from the group consisting of

i. A112E/T226A;
ii. S196T/S411F;
iii. E295K/S398T;
iv. T246C/Y371C;
v. T41I plus deletion at T445
vi. A68T/G440R/P491L;
vii. G22D/S278P/T296P;
viii. T246A/R251A/Y252A;
ix. T380G/Y381D/R394A;
x. T380G/Y381D/R394A plus deletion of 382-393, inclusive;
xi. Y252Q/D259W/S342Y;
xii. S113T/T255P/K286M;
xiii. P227L/E325K/Q487L;
xiv. P227T/T484S/F352L;
xv. Q17L/E193V/M213I/F352L;
xvi. S8P/N49S/A68T/S113N;
xvii. S8P/N49S/A68T/S113N/P227L;
xviii. T41I/A112E/P227L/S278P/T296P;
xix. S8P/N49S/A68T/A112E/T226A;
xx. S8P/N49S/A68T/A112E/P227L;
xxi. S8P/T41I/N49S/A68T/A112E/P227L;

xxii. G22D/N49S/A68T/P227L/S278P/T296P;

xxiii. S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/S278P/T296P;

xxiv. G22D/N49S/A68T/N103I/S113N/P227L/S278P/T296P;

xxv. G22D/N49S/A68T/N103I/A112E/P227L/S278P/T296P;

xxvi. G22D/N49S/N64D/A68T/N103I/S113N/S278P/T296P;

xxvii. S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/D249K/S278P/T296P;

xxviii. S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/S278P/T296P/N301R;

xxix. S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/D249K/S278P/T296P/N301R

xxx. S8P/G22D/T41I/N49S/A68T/S113N/P227L/D249K/S278P/T296P/N301R;

xxxi. S8P/T41I/N49S/S57N/A68T/S113N/P227L/D249K/S278P/T296P/N301R;

xxxii. S8P/G22D/T41I/N49S/A68T/S113N/P227L/D249K/S278P/N301R;

xxxiii. S8P/T41I/N49S/A68T/S92T/S113N/P227L/D249K/V403D/T462I;

xxxiv. S8P/G22D/T41I/N49S/A68T/S92T/S113N/P227L/D249K/V403D/T462I;

xxxv. S8P/T41I/N49S/A68T/S92T/S113N/P227L/D249K/S411F;

xxxvi. S8P/G22D/T41I/N49S/A68T/S92T/S113N/P227L/D249K/S411F;

xxxvii. S8P/G22D/T41I/N49S/A68T/S92T/S113N/S196T/P227L/D249K/T255P/ S278P/T296P/N301R/E325K/S411F;

xxxviii. S8P/T41I/N49S/A68T/S92T/S113N/S196T/P227L/D249K/T255P/S278P/T296P/N301R/E325K/V403D/S411F/T462I;

xxxix. S8P/G22D/T41I/N49S/A68T/S92T/S113N/S196T/P227L/D249K/T255P/ S278P/T296P/N301R/E325K/V403D/S411F/T462I;

in CBH I from *Hypocrea jecorina* (SEQ ID NO:2).

[18] In an fifth embodiment the invention is directed to a vector comprising a nucleic acid encoding a variant CBH I. In another aspect there is a construct comprising the nucleic acid of encoding the variant CBH I operably linked to a regulatory sequence.

[19] In a sixth embodiment the invention is directed to a host cell transformed with the vector comprising a nucleic acid encoding a CBH I variant.

[20] In a seventh embodiment the invention is directed to a method of producing a CBH I variant comprising the steps of:

(a) culturing a host cell transformed with the vector comprising a nucleic acid encoding a CBH I variant in a suitable culture medium under suitable conditions to produce CBH I variant;

(b) obtaining said produced CBH I variant.

[21] In an eighth embodiment the invention is directed to a detergent composition comprising a surfactant and a CBH I variant. In one aspect of this embodiment the detergent is a laundry detergent. In a second aspect of this embodiment the detergent is a dish detergent. In third aspect of this invention, the variant CBH I cellulase is used in the treatment of a cellulose containing textile, in particular, in the stonewashing or indigo dyed denim.

[22] In a ninth embodiment the invention is directed to a feed additive comprising a CBH I variant.

[23] In a tenth embodiment the invention is directed to a method of treating wood pulp comprising contacting said wood pulp with a CBH I variant.

[24] In a eleventh embodiment the invention is directed to a method of converting biomass to sugars comprising contacting said biomass with a CBH I variant.

[25] In an embodiment, the cellulase is derived from a fungus, bacteria or Actinomycete. In another aspect, the cellulase is derived from a fungus. In a most preferred embodiment, the fungus is a filamentous fungus. It is preferred the filamentous fungus belong to Euascomycete, in particular, *Aspergillus spp., Gliocladium spp., Fusarium spp., Acremonium spp., Myceliophtora spp., Verticillium spp., Myrothecium spp.,* or *Penicillium spp.* In a further aspect of this embodiment, the cellulase is a cellobiohydrolase.

[26] Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the scope and spirit of the invention will become apparent to one skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[27] Figure 1 is the nucleic acid (lower line; SEQ ID NO: 1) and amino acid (upper line; SEQ ID NO: 2) sequence of the wild type Cel7A (CBH I) from *H. jecorina*.

[28] Figure 2 is the 3-D structure of *H. jecorina* CBH I.

[29] Figure 3 shows the amino acid alignment of the Cel7 family members for which there were crystal structures available. The sequences are: 2OVW - Fusarium oxysporum Cel7B, 1A39 - Humicola insolens Cel7B, 6CEL - Hypocrea jecorina Cel7A, 1EG1 - Hypocrea jecorina Cel7B.

[30] Figure 4 illustrates the crystal structures from the catalytic domains of these four Cel7 homologues aligned and overlayed as described herein.

[31] Figure 5 A-M is the nucleic acid sequence and deduced amino acid sequence for eight single residue mutations and five multiple mutation variants.

[32] Figure 6 A-D is the nucleic acid sequence for pTrex2.

[33] Figure 7 A & B depicts the construction of the expression plasmid pTEX.

[34] Figure 8 A-J is the amino acid alignment of all 42 members of the Cel7 family.

[35] Figure 9A is a representation of the thermal profiles of the wild type and eight single residue variants. Figure 9B is a representation of the thermal profiles of the wild type and five variants. Legend for Figure 9B: Cel7A = wild-type H. jecorina CBH I; N301K = N301K variant; 334 = P227L variant; 340 = S8P/N49S/A68T/S113N variant; 350 = S8P/N49S/A68T/S113N/ P227L variant; and 363 = S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/S278P/T296P variant.

[36] Figure 10 is the pRAX1 vector. This vector is based on the plasmid pGAPT2 except a 5259bp HindIII fragment of Aspergillus nidulans genomic DNA fragment AMA1 sequence (Molecular Microbiology 1996 19:565-574) was inserted. Base 1 to 1134 contains Aspergillus niger glucoamylase gene promoter. Base 3098 to 3356 and 4950 to 4971 contains Aspergillus niger glucoamylase terminator. Aspergillus nidulans pyrG gene was inserted from 3357 to 4949 as a marker for fungal transformation. There is a multiple cloning site (MCS) into which genes may be inserted.

[37] Figure 11 is the pRAXdes2 vector backbone. This vector is based on the plasmid vector pRAX1. A Gateway cassette has been inserted into pRAX1 vector (indicated by the arrow on the interior of the circular plasmid). This cassette contains recombination sequence attR1 and attR2 and the selection marker catH and ccdB. The vector has been made according to the manual given in Gateway™ Cloning Technology: version 1 page 34-38 and can only replicate in *E. coli* DB3.1 from Invitrogen; in other *E. coli* hosts the ccdB gene is lethal. First a PCR fragment is made with primers containing attB1/2 recombination sequences. This fragment is recombined with pDONR201 (commercially available from Invitrogen); this vector contains attP1/2 recombination sequences with catH and ccdB in between the recombination sites. The BP clonase enzymes from Invitrogen are used to recombine the PCR fragment in this so-called ENTRY vector, clones with the

PCR fragment inserted can be selected at 50µg/ml kanamycin because clones expressing ccdB do not survive. Now the att sequences are altered and called attL1 and attL2. The second step is to recombine this clone with the pRAXdes2 vector (containing attR1 and attR2 catH and ccdB in between the recombination sites). The LR clonase enzymes from Invitrogen are used to recombine the insert from the ENTRY vector in the destination vector. Only pRAXCBH1 vectors are selected using 100µg/ml ampicillin because ccdB is lethal and the ENTRY vector is sensitive to ampicillin. By this method the expression vector is now prepared and can be used to transform *A. niger*.

[38] Figure 12 provides an illustration of the pRAXdes2cbh1 vector which was used for expression of the nucleic acids encoding the CBH1 variants in Aspergillus. A nucleic acid encoding a CBH1 enzyme homolog or variant was cloned into the vector by homologous recombination of the att sequences.

DETAILED DESCRIPTION

[39] The invention will now be described in detail by way of reference only using the following definitions and examples. All patents and publications, including all sequences disclosed within such patents and publications, referred to herein are expressly incorporated by reference.

[40] Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton, *et al.,* DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2D ED., John Wiley and Sons, New York (1994), and Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) provide one of skill with a general dictionary of many of the terms used in this invention. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively. Practitioners are particularly directed to Sambrook *et al.,* MOLECULAR CLONING: A LABORATORY MANUAL (Second Edition), Cold Spring Harbor Press, Plainview, N.Y., 1989, and Ausubel FM *et al.,* Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, 1993, for definitions and terms of the art. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary.

[41] The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

[42] All publications cited herein are expressly incorporated herein by reference for the purpose of describing and disclosing compositions and methodologies which might be used in connection with the invention.

I. DEFINITIONS

[43] The term "polypeptide" as used herein refers to a compound made up of a single chain of amino acid residues linked by peptide bonds. The term "protein" as used herein may be synonymous with the term "polypeptide" or may refer, in addition, to a complex of two or more polypeptides.

[44] "Variant" means a protein which is derived from a precursor protein (*e.g.*, the native protein) by addition of one or more amino acids to either or both the C- and N-terminal end, substitution of one or more amino acids at one or a number of different sites in the amino acid sequence, or deletion of one or more amino acids at either or both ends of the protein or at one or more sites in the amino acid sequence. The preparation of an enzyme variant is preferably achieved by modifying a DNA sequence which encodes for the native protein, transformation of that DNA sequence into a suitable host, and expression of the modified DNA sequence to form the derivative enzyme. The variant CBH I enzyme of the invention includes peptides comprising altered amino acid sequences in comparison with a precursor enzyme amino acid sequence wherein the variant CBH enzyme retains the characteristic cellulolytic nature of the precursor enzyme but which may have altered properties in some specific aspect. For example, a variant CBH enzyme may have an increased pH optimum or increased temperature or oxidative stability but will retain its characteristic cellulolytic activity. It is contemplated that the variants according to the present invention may be derived from a DNA fragment encoding a cellulase variant CBH enzyme wherein the functional activity of the expressed cellulase derivative is retained. For example, a DNA fragment encoding a cellulase may further include a DNA sequence or portion thereof encoding a hinge or linker attached to the cellulase DNA sequence at either the 5' or 3' end wherein the functional activity of the encoded cellulase domain is retained.

[45] "Equivalent residues" may also be defined by determining homology at the level of tertiary structure for a precursor cellulase whose tertiary structure has been determined by x-ray crystallography. Equivalent residues are defined as those for which the atomic

coordinates of two or more of the main chain atoms of a particular amino acid residue of a cellulase and *Hypocrea jecorina* CBH (N on N, CA on CA, C on C and O on O) are within 0.13nm and preferably 0.1nm after alignment. Alignment is achieved after the best model has been oriented and positioned to give the maximum overlap of atomic coordinates of non-hydrogen protein atoms of the cellulase in question to the *H. jecorina* CBH I. The best model is the crystallographic model giving the lowest R factor for experimental diffraction data at the highest resolution available.

$$R\,factor = \frac{\Sigma_h |Fo(h)| - |Fc(h)|}{\Sigma_h |Fo(h)|}$$

[46] Equivalent residues which are functionally analogous to a specific residue of *H. jecorina* CBH I are defined as those amino acids of a cellulase which may adopt a conformation such that they either alter, modify or contribute to protein structure, substrate binding or catalysis in a manner defined and attributed to a specific residue of the *H. jecorina* CBH I. Further, they are those residues of the cellulase (for which a tertiary structure has been obtained by x-ray crystallography) which occupy an analogous position to the extent that, although the main chain atoms of the given residue may not satisfy the criteria of equivalence on the basis of occupying a homologous position, the atomic coordinates of at least two of the side chain atoms of the residue lie with 0.13nm of the corresponding side chain atoms of *H. jecorina* CBH. The crystal structure of *H. jecorina* CBH I is shown in Figure 2.

[47] The term "nucleic acid molecule" includes RNA, DNA and cDNA molecules. It will be understood that, as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding a given protein such as CBH I may be produced. The present invention contemplates every possible variant nucleotide sequence, encoding CBH I, all of which are possible given the degeneracy of the genetic code.

[48] A "heterologous" nucleic acid construct or sequence has a portion of the sequence which is not native to the cell in which it is expressed. Heterologous, with respect to a control sequence refers to a control sequence (*i.e.* promoter or enhancer) that does not function in nature to regulate the same gene the expression of which it is currently regulating. Generally, heterologous nucleic acid sequences are not endogenous to the cell or part of the genome in which they are present, and have been added to the cell, by infection, transfection, transformation, microinjection, electroporation, or the like. A "heterologous" nucleic acid construct may contain a control sequence/DNA coding

sequence combination that is the same as, or different from a control sequence/DNA coding sequence combination found in the native cell.

[49] As used herein, the term "vector" refers to a nucleic acid construct designed for transfer between different host cells. An "expression vector" refers to a vector that has the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many prokaryotic and eukaryotic expression vectors are commercially available. Selection of appropriate expression vectors is within the knowledge of those having skill in the art.

[50] Accordingly, an “expression cassette” or “expression vector” is a nucleic acid construct generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a target cell. The recombinant expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus, or nucleic acid fragment. Typically, the recombinant expression cassette portion of an expression vector includes, among other sequences, a nucleic acid sequence to be transcribed and a promoter.

[51] As used herein, the term "plasmid" refers to a circular double-stranded (ds) DNA construct used as a cloning vector, and which forms an extrachromosomal self-replicating genetic element in many bacteria and some eukaryotes.

[52] As used herein, the term "selectable marker-encoding nucleotide sequence" refers to a nucleotide sequence which is capable of expression in cells and where expression of the selectable marker confers to cells containing the expressed gene the ability to grow in the presence of a corresponding selective agent, or under corresponding selective growth conditions.

[53] As used herein, the term "promoter" refers to a nucleic acid sequence that functions to direct transcription of a downstream gene. The promoter will generally be appropriate to the host cell in which the target gene is being expressed. The promoter together with other transcriptional and translational regulatory nucleic acid sequences (also termed "control sequences") are necessary to express a given gene. In general, the transcriptional and translational regulatory sequences include, but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences.

[54] "Chimeric gene" or “heterologous nucleic acid construct”, as defined herein refers to a non-native gene (*i.e.*, one that has been introduced into a host) that may be composed of parts of different genes, including regulatory elements. A chimeric gene construct for transformation of a host cell is typically composed of a transcriptional regulatory region (promoter) operably linked to a heterologous protein coding sequence,

or, in a selectable marker chimeric gene, to a selectable marker gene encoding a protein conferring antibiotic resistance to transformed cells. A typical chimeric gene of the present invention, for transformation into a host cell, includes a transcriptional regulatory region that is constitutive or inducible, a protein coding sequence, and a terminator sequence. A chimeric gene construct may also include a second DNA sequence encoding a signal peptide if secretion of the target protein is desired.

[55] A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA encoding a secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading frame. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors, linkers or primers for PCR are used in accordance with conventional practice.

[56] As used herein, the term "gene" means the segment of DNA involved in producing a polypeptide chain, that may or may not include regions preceding and following the coding region, *e.g.* 5' untranslated (5' UTR) or "leader" sequences and 3' UTR or "trailer" sequences, as well as intervening sequences (introns) between individual coding segments (exons).

[57] In general, nucleic acid molecules which encode the variant CBH I will hybridize, under moderate to high stringency conditions to the wild type sequence provided herein as SEQ ID NO:1. However, in some cases a CBH I-encoding nucleotide sequence is employed that possesses a substantially different codon usage, while the protein encoded by the CBH I-encoding nucleotide sequence has the same or substantially the same amino acid sequence as the native protein. For example, the coding sequence may be modified to facilitate faster expression of CBH I in a particular prokaryotic or eukaryotic expression system, in accordance with the frequency with which a particular codon is utilized by the host. Te'o, *et al.* (FEMS Microbiology Letters 190:13-19, 2000), for example, describes the optimization of genes for expression in filamentous fungi.

[58] A nucleic acid sequence is considered to be "selectively hybridizable" to a reference nucleic acid sequence if the two sequences specifically hybridize to one another under moderate to high stringency hybridization and wash conditions. Hybridization

conditions are based on the melting temperature (Tm) of the nucleic acid binding complex or probe. For example, "maximum stringency" typically occurs at about Tm-5°C (5° below the Tm of the probe); "high stringency" at about 5-10° below the Tm; "moderate " or "intermediate stringency" at about 10-20° below the Tm of the probe; and "low stringency" at about 20-25° below the Tm. Functionally, maximum stringency conditions may be used to identify sequences having strict identity or near-strict identity with the hybridization probe; while high stringency conditions are used to identify sequences having about 80% or more sequence identity with the probe.

[59] Moderate and high stringency hybridization conditions are well known in the art (see, for example, Sambrook, *et al,* 1989, Chapters 9 and 11, and in Ausubel, F.M., *et al.*, 1993, expressly incorporated by reference herein). An example of high stringency conditions includes hybridization at about 42°C in 50% formamide, 5X SSC, 5X Denhardt's solution, 0.5% SDS and 100 μg/ml denatured carrier DNA followed by washing two times in 2X SSC and 0.5% SDS at room temperature and two additional times in 0.1X SSC and 0.5% SDS at 42°C.

[60] As used herein, "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid sequence or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all as a result of deliberate human intervention.

[61] As used herein, the terms "transformed", "stably transformed" or "transgenic" with reference to a cell means the cell has a non-native (heterologous) nucleic acid sequence integrated into its genome or as an episomal plasmid that is maintained through multiple generations.

[62] As used herein, the term "expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation.

[63] The term "introduced" in the context of inserting a nucleic acid sequence into a cell, means "transfection", or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid sequence into a eukaryotic or prokaryotic cell where the nucleic acid sequence may be incorporated into the genome of the cell (for example, chromosome, plasmid, plastid, or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (for example, transfected mRNA).

[64] It follows that the term "CBH I expression" refers to transcription and translation of the *cbh I* gene, the products of which include precursor RNA, mRNA, polypeptide, post-translationally processed polypeptides, and derivatives thereof, including CBH I from related species such as *Trichoderma koningii, Hypocrea jecorina (also known as Trichoderma longibrachiatum, Trichoderma reesei or Trichoderma viride) and Hypocrea schweinitzii*. By way of example, assays for CBH I expression include Western blot for CBH I protein, Northern blot analysis and reverse transcriptase polymerase chain reaction (RT-PCR) assays for CBH I mRNA, and *en*doglucanase activity assays as described in Shoemaker S.P. and Brown R.D.Jr. (Biochim. Biophys. Acta, 1978, 523:133-146) and Schulein (Methods Enzymol., 160, 25, pp. 234-243, 1988).

[65] The term "alternative splicing" refers to the process whereby multiple polypeptide isoforms are generated from a single gene, and involves the splicing together of nonconsecutive exons during the processing of some, but not all, transcripts of the gene. Thus a particular exon may be connected to any one of several alternative exons to form messenger RNAs. The alternatively-spliced mRNAs produce polypeptides ("splice variants") in which some parts are common while other parts are different.

[66] The term "signal sequence" refers to a sequence of amino acids at the N-terminal portion of a protein which facilitates the secretion of the mature form of the protein outside the cell. The mature form of the extracellular protein lacks the signal sequence which is cleaved off during the secretion process.

[67] By the term "host cell" is meant a cell that contains a vector and supports the replication, and/or transcription or transcription and translation (expression) of the expression construct. Host cells for use in the present invention can be prokaryotic cells, such as *E. coli*, or eukaryotic cells such as yeast, plant, insect, amphibian, or mammalian cells. In general, host cells are filamentous fungi.

[68] The term "filamentous fungi" means any and all filamentous fungi recognized by those of skill in the art. A preferred fungus is selected from the group consisting of Aspergillus, Trichoderma, Fusarium, Chrysosporium, Penicillium, Humicola, Neurospora, or alternative sexual forms thereof such as Emericella, Hypocrea. It has now been demonstrated that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina*. See Kuhls et al., PNAS (1996) 93:7755-7760.

[69] The term "cellooligosaccharide" refers to oligosaccharide groups containing from 2-8 glucose units and having β-1,4 linkages, *e.g.*, cellobiose.

[70] The term "cellulase" refers to a category of enzymes capable of hydrolyzing cellulose polymers to shorter cello-oligosaccharide oligomers, cellobiose and/or glucose.

Numerous examples of cellulases, such as exoglucanases, exocellobiohydrolases, endoglucanases, and glucosidases have been obtained from cellulolytic organisms, particularly including fungi, plants and bacteria.

[71] CBH I from *Hypocrea jecorina* is a member of the Glycosyl Hydrolase Family 7 (hence Cel 7) and, specifically, was the first member of that family identified in *Hypocrea jecorina* (hence Cel 7A). The Glycosyl Hydrolase Family 7 contains both Endoglucanases and Cellobiohydrolases/exoglucanases, and that CBH I is the latter. Thus, the phrases CBH I, CBH I-type protein and Cel 7 cellobiohydrolases may be used interchangeably herein.

[72] The term "cellulose binding domain" as used herein refers to portion of the amino acid sequence of a cellulase or a region of the enzyme that is involved in the cellulose binding activity of a cellulase or derivative thereof. Cellulose binding domains generally function by non-covalently binding the cellulase to cellulose, a cellulose derivative or other polysaccharide equivalent thereof. Cellulose binding domains permit or facilitate hydrolysis of cellulose fibers by the structurally distinct catalytic core region, and typically function independent of the catalytic core. Thus, a cellulose binding domain will not possess the significant hydrolytic activity attributable to a catalytic core. In other words, a cellulose binding domain is a structural element of the cellulase enzyme protein tertiary structure that is distinct from the structural element which possesses catalytic activity. Cellulose binding domain and cellulose binding module may be used interchangeably herein.

[73] As used herein, the term "surfactant" refers to any compound generally recognized in the art as having surface active qualities. Thus, for example, surfactants comprise anionic, cationic and nonionic surfactants such as those commonly found in detergents. Anionic surfactants include linear or branched alkylbenzenesulfonates; alkyl or alkenyl ether sulfates having linear or branched alkyl groups or alkenyl groups; alkyl or alkenyl sulfates; olefinsulfonates; and alkanesulfonates. Ampholytic surfactants include quaternary ammonium salt sulfonates, and betaine-type ampholytic surfactants. Such ampholytic surfactants have both the positive and negative charged groups in the same molecule. Nonionic surfactants may comprise polyoxyalkylene ethers, as well as higher fatty acid alkanolamides or alkylene oxide adduct thereof, fatty acid glycerine monoesters, and the like.

[74] As used herein, the term "cellulose containing fabric" refers to any sewn or unsewn fabrics, yarns or fibers made of cotton or non-cotton containing cellulose or cotton or non-

cotton containing cellulose blends including natural cellulosics and manmade cellulosics (such as jute, flax, ramie, rayon, and lyocell).

[75] As used herein, the term "cotton-containing fabric" refers to sewn or unsewn fabrics, yarns or fibers made of pure cotton or cotton blends including cotton woven fabrics, cotton knits, cotton denims, cotton yarns, raw cotton and the like.

[76] As used herein, the term "stonewashing composition" refers to a formulation for use in stonewashing cellulose containing fabrics. Stonewashing compositions are used to modify cellulose containing fabrics prior to sale, *i.e.*, during the manufacturing process. In contrast, detergent compositions are intended for the cleaning of soiled garments and are not used during the manufacturing process.

[77] As used herein, the term "detergent composition" refers to a mixture which is intended for use in a wash medium for the laundering of soiled cellulose containing fabrics. In the context of the present invention, such compositions may include, in addition to cellulases and surfactants, additional hydrolytic enzymes, builders, bleaching agents, bleach activators, bluing agents and fluorescent dyes, caking inhibitors, masking agents, cellulase activators, antioxidants, and solubilizers.

[78] As used herein, the term "decrease or elimination in expression of the *cbh1* gene" means that either that the *cbh1* gene has been deleted from the genome and therefore cannot be expressed by the recombinant host microorganism; or that the *cbh1* gene has been modified such that a functional CBH1 enzyme is not produced by the host microorganism.

[79] The term "variant *cbh1 gene*" or "variant CBH1" means, respectively, that the nucleic acid sequence of the *cbh1* gene from *H. jecorina* has been altered by removing, adding, and/or manipulating the coding sequence or the amino acid sequence of the expressed protein has been modified consistent with the invention described herein.

[80] As used herein, the term "purifying" generally refers to subjecting transgenic nucleic acid or protein containing cells to biochemical purification and/or column chromatography.

[81] As used herein, the terms "active" and "biologically active" refer to a biological activity associated with a particular protein and are used interchangeably herein. For example, the enzymatic activity associated with a protease is proteolysis and, thus, an active protease has proteolytic activity. It follows that the biological activity of a given protein refers to any biological activity typically attributed to that protein by those of skill in the art.

[82] As used herein, the term "enriched" means that the CBH is found in a concentration that is greater relative to the CBH concentration found in a wild-type, or naturally occurring, fungal cellulase composition. The terms enriched, elevated and enhanced may be used interchangeably herein.

[83] A wild type fungal cellulase composition is one produced by a naturally occurring fungal source and which comprises one or more BGL, CBH and EG components wherein each of these components is found at the ratio produced by the fungal source. Thus, an enriched CBH composition would have CBH at an altered ratio wherein the ratio of CBH to other cellulase components (i.e., EGs, beta-glucosidases and other endoglucanases) is elevated. This ratio may be increased by either increasing CBH or decreasing (or eliminating) at least one other component by any means known in the art.

[84] Thus, to illustrate, a naturally occurring cellulase system may be purified into substantially pure components by recognized separation techniques well published in the literature, including ion exchange chromatography at a suitable pH, affinity chromatography, size exclusion and the like. For example, in ion exchange chromatography (usually anion exchange chromatography), it is possible to separate the cellulase components by eluting with a pH gradient, or a salt gradient, or both a pH and a salt gradient. The purified CBH may then be added to the enzymatic solution resulting in an enriched CBH solution. It is also possible to elevate the amount of CBH I produced by a microbe using molecular genetics methods to overexpress the gene encoding CBH, possibly in conjunction with deletion of one or more genes encoding other cellulases.

[85] Fungal cellulases may contain more than one CBH component. The different components generally have different isoelectric points which allow for their separation via ion exchange chromatography and the like. Either a single CBH component or a combination of CBH components may be employed in an enzymatic solution.

[86] When employed in enzymatic solutions, the homolog or variant CBH1 component is generally added in an amount sufficient to allow the highest rate of release of soluble sugars from the biomass. The amount of homolog or variant CBH1 component added depends upon the type of biomass to be saccharified which can be readily determined by the skilled artisan. However, when employed, the weight percent of the homolog or variant CBH1 component relative to any EG type components present in the cellulase composition is from preferably about 1, preferably about 5, preferably about 10, preferably about 15, or preferably about 20 weight percent to preferably about 25, preferably about 30, preferably about 35, preferably about 40, preferably about 45 or preferably about 50 weight percent. Furthermore, preferred ranges may be about 0.5 to about 15 weight

percent, about 0.5 to about 20 weight percent, from about 1 to about 10 weight percent, from about 1 to about 15 weight percent, from about 1 to about 20 weight percent, from about 1 to about 25 weight percent, from about 5 to about 20 weight percent, from about 5 to about 25 weight percent, from about 5 to about 30 weight percent, from about 5 to about 35 weight percent, from about 5 to about 40 weight percent, from about 5 to about 45 weight percent, from about 5 to about 50 weight percent, from about 10 to about 20 weight percent, from about 10 to about 25 weight percent, from about 10 to about 30 weight percent, from about 10 to about 35 weight percent, from about 10 to about 40 weight percent, from about 10 to about 45 weight percent, from about 10 to about 50 weight percent, from about 15 to about 20 weight percent, from about 15 to about 25 weight percent, from about 15 to about 30 weight percent, from about 15 to about 35 weight percent, from about 15 to about 30 weight percent, from about 15 to about 45 weight percent, from about 15 to about 50 weight percent.

II. HOST ORGANISMS

[87] Filamentous fungi include all filamentous forms of the subdivision Eumycota and Oomycota. The filamentous fungi are characterized by vegetative mycelium having a cell wall composed of chitin, glucan, chitosan, mannan, and other complex polysaccharides, with vegetative growth by hyphal elongation and carbon catabolism that is obligately aerobic.

[88] In the present invention, the filamentous fungal parent cell may be a cell of a species of, but not limited to, *Trichoderma, e.g., Trichoderma longibrachiatum, Trichoderma viride, Trichoderma koningii, Trichoderma harzianum; Penicillium sp.; Humicola sp.*, including *Humicola insolens* and *Humicola grisea; Chrysosporium sp.*, including *C. lucknowense*; *Gliocladium sp.; Aspergillus sp.*; *Fusarium sp., Neurospora sp., Hypocrea sp., and Emericella sp.* As used herein, the term "*Trichoderma*" or "*Trichoderma sp.*" refers to any fungal strains which have previously been classified as *Trichoderma* or are currently classified as *Trichoderma.*

[89] In one preferred embodiment, the filamentous fungal parent cell is an *Aspergillus niger, Aspergillus awamori, Aspergillus aculeatus,* or *Aspergillus nidulans* cell.

[90] In another preferred embodiment, the filamentous fungal parent cell is a *Trichoderma reesei* cell.

III. CELLULASES

[91] Cellulases are known in the art as enzymes that hydrolyze cellulose (beta-1,4-glucan or beta D-glucosidic linkages) resulting in the formation of glucose, cellobiose,

cellooligosaccharides, and the like. As set forth above, cellulases have been traditionally divided into three major classes: endoglucanases (EC 3.2.1.4) ("EG"), exoglucanases or cellobiohydrolases (EC 3.2.1.91) ("CBH") and beta-glucosidases (EC 3.2.1.21) ("BG"). (Knowles, *et al.*, TIBTECH 5, 255-261, 1987; Schulein, 1988).

[92] Certain fungi produce complete cellulase systems which include exo-cellobiohydrolases or CBH-type cellulases, endoglucanases or EG-type cellulases and beta-glucosidases or BG-type cellulases (Schulein, 1988). However, sometimes these systems lack CBH-type cellulases and bacterial cellulases also typically include little or no CBH-type cellulases. In addition, it has been shown that the EG components and CBH components synergistically interact to more efficiently degrade cellulose. See, *e.g.*, Wood, 1985. The different components, *i.e.*, the various endoglucanases and exocellobiohydrolases in a multi-component or complete cellulase system, generally have different properties, such as isoelectric point, molecular weight, degree of glycosylation, substrate specificity and enzymatic action patterns.

[93] It is believed that endoglucanase-type cellulases hydrolyze internal beta -1,4-glucosidic bonds in regions of low crystallinity of the cellulose and exo-cellobiohydrolase-type cellulases hydrolyze cellobiose from the reducing or non-reducing end of cellulose. It follows that the action of endoglucanase components can greatly facilitate the action of exo-cellobiohydrolases by creating new chain ends which are recognized by exo-cellobiohydrolase components. Further, beta-glucosidase-type cellulases have been shown to catalyze the hydrolysis of alkyl and/or aryl β-D-glucosides such as methyl β-D-glucoside and p-nitrophenyl glucoside as well as glycosides containing only carbohydrate residues, such as cellobiose. This yields glucose as the sole product for the microorganism and reduces or eliminates cellobiose which inhibits cellobiohydrolases and endoglucanases.

[94] Cellulases also find a number of uses in detergent compositions including to enhance cleaning ability, as a softening agent and to improve the feel of cotton fabrics (Hemmpel, ITB Dyeing/Printing/Finishing 3:5-14, 1991; Tyndall, Textile Chemist and Colorist 24:23-26, 1992; Kumar *et al.*, Textile Chemist and Colorist, 29:37-42, 1997). While the mechanism is not part of the invention, softening and color restoration properties of cellulase have been attributed to the alkaline endoglucanase components in cellulase compositions, as exemplified by U.S. Patent Nos. 5,648,263, 5,691,178, and 5,776,757, which disclose that detergent compositions containing a cellulase composition enriched in a specified alkaline endoglucanase component impart color restoration and improved softening to treated garments as compared to cellulase compositions not enriched in such

a component. In addition, the use of such alkaline endoglucanase components in detergent compositions has been shown to complement the pH requirements of the detergent composition (*e.g.*, by exhibiting maximal activity at an alkaline pH of 7.5 to 10, as described in U.S. Patent Nos. 5,648,263, 5,691,178, and 5,776,757).

[95] Cellulase compositions have also been shown to degrade cotton-containing fabrics, resulting in reduced strength loss in the fabric (U.S. Patent No. 4,822,516), contributing to reluctance to use cellulase compositions in commercial detergent applications. Cellulase compositions comprising endoglucanase components have been suggested to exhibit reduced strength loss for cotton-containing fabrics as compared to compositions comprising a complete cellulase system.

[96] Cellulases have also been shown to be useful in degradation of cellulase biomass to ethanol (wherein the cellulase degrades cellulose to glucose and yeast or other microbes further ferment the glucose into ethanol), in the treatment of mechanical pulp (Pere *et al.*, 1996), for use as a feed additive (WO 91/04673) and in grain wet milling.

[97] Most CBHs and EGs have a multidomain structure consisting of a core domain separated from a cellulose binding domain (CBD) by a linker peptide (Suurnakki *et al.*, 2000). The core domain contains the active site whereas the CBD interacts with cellulose by binding the enzyme to it (van Tilbeurgh *et al.*, 1986; Tomme *et al.*, Eur. J. Biochem. 170:575-581, 1988). The CBDs are particularly important in the hydrolysis of crystalline cellulose. It has been shown that the ability of cellobiohydrolases to degrade crystalline cellulose clearly decreases when the CBD is absent (Linder and Teeri, J. Biotechnol. 57:15-28, 1997). However, the exact role and action mechanism of CBDs is still a matter of speculation. It has been suggested that the CBD enhances the enzymatic activity merely by increasing the effective enzyme concentration at the surface of cellulose (Stahlberg *et al.*, Bio/Technol. 9:286-290, 1991), and/or by loosening single cellulose chains from the cellulose surface (Tormo *et al.*, EMBO J. vol. 15, no. 21, pp. 5739-5751, 1996). Most studies concerning the effects of cellulase domains on different substrates have been carried out with core proteins of cellobiohydrolases, as their core proteins can easily be produced by limited proteolysis with papain (Tomme *et al.*, 1988). Numerous cellulases have been described in the scientific literature, examples of which include: from *Trichoderma reesei*: Shoemaker, S. et al., Bio/Technology, 1:691-696, 1983, which discloses *CBHI*; Teeri, T. et al., Gene, 51:43-52, 1987, which discloses *CBHII*. Cellulases from species other than Trichoderma have also been described e.g., Ooi *et al.*, Nucleic Acids Research, vol. 18, no. 19, 1990, which discloses the cDNA sequence coding for endoglucanase F1-CMC produced by *Aspergillus aculeatus*; Kawaguchi T *et al.*, Gene

173(2):287-8, 1996, which discloses the cloning and sequencing of the cDNA encoding beta-glucosidase 1 from *Aspergillus aculeatus*; Sakamoto *et al.*, Curr. Genet. 27:435-439, 1995, which discloses the cDNA sequence encoding the endoglucanase CMCase-1 from *Aspergillus kawachii* IFO 4308; Saarilahti *et al.*, Gene 90:9-14, 1990, which discloses an endoglucanase from *Erwinia carotovara*; Spilliaert R, *et al.*, Eur J Biochem. 224(3):923-30, 1994, which discloses the cloning and sequencing of bglA, coding for a thermostable beta-glucanase from *Rhodothermus marinu*; and Halldorsdottir S *et al.*, Appl Microbiol Biotechnol. 49(3):277-84, 1998, which discloses the cloning, sequencing and overexpression of a *Rhodothermus marinus* gene encoding a thermostable cellulase of glycosyl hydrolase family 12. However, there remains a need for identification and characterization of novel cellulases, with improved properties, such as improved performance under conditions of thermal stress or in the presence of surfactants, increased specific activity, altered substrate cleavage pattern, and/or high level expression *in vitro*.

[98] The development of new and improved cellulase compositions that comprise varying amounts CBH-type, EG-type and BG-type cellulases is of interest for use: (1) in detergent compositions that exhibit enhanced cleaning ability, function as a softening agent and/or improve the feel of cotton fabrics (*e.g.*, "stone washing" or "biopolishing"); (2) in compositions for degrading wood pulp or other biomass into sugars (*e.g.*, for bio-ethanol production); and/or (3) in feed compositions.

IV. MOLECULAR BIOLOGY

[99] In one embodiment this invention provides for the expression of variant CBH I genes under control of a promoter functional in a filamentous fungus. Therefore, this invention relies on routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and Ausubel *et al.*, eds., *Current Protocols in Molecular Biology* (1994)).

A. Methods for Identifying Homologous CBH1 Genes

[100] The nucleic acid sequence for the wild type *H. jecorina* CBH1 is shown in Figure 1. The invention, in one aspect, encompasses a nucleic acid molecule encoding a CBH1 homolog described herein. The nucleic acid may be a DNA molecule.

[101] Techniques that can be used to isolate CBH I encoding DNA sequences are well known in the art and include, but are not limited to, cDNA and/or genomic library screening with a homologous DNA probe and expression screening with activity assays or antibodies against CBH I. Any of these methods can be found in Sambrook, *et al.* or in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, F. Ausubel, *et al.*, ed. Greene Publishing and Wiley-Interscience, New York (1987) ("Ausubel").

B. Methods of Mutating CBH I Nucleic Acid Sequences

[102] Any method known in the art that can introduce mutations is contemplated by the present invention.

[103] The present invention relates to the expression, purification and/or isolation and use of variant CBH1. These enzymes are preferably prepared by recombinant methods utilizing the *cbh* gene from *H. jecorina*.

[104] After the isolation and cloning of the *cbh1* gene from *H. jecorina*, other methods known in the art, such as site directed mutagenesis, are used to make the substitutions, additions or deletions that correspond to substituted amino acids in the expressed CBH1 variant. Again, site directed mutagenesis and other methods of incorporating amino acid changes in expressed proteins at the DNA level can be found in Sambrook, *et al.* and Ausubel, *et al.*

[105] DNA encoding an amino acid sequence variant of the *H. jecorina* CBH1 is prepared by a variety of methods known in the art. These methods include, but are not limited to, preparation by site-directed (or oligonucleotide-mediated) mutagenesis, PCR mutagenesis, and cassette mutagenesis of an earlier prepared DNA encoding the *H. jecorina* CBH1.

[106] Site-directed mutagenesis is a preferred method for preparing substitution variants. This technique is well known in the art (see, e.g.,Carter et al. Nucleic Acids Res. 13:4431-4443 (1985) and Kunkel et al., Proc. Natl. Acad.Sci.USA 82:488 (1987)). Briefly, in carrying out site-directed mutagenesis of DNA, the starting DNA is altered by first hybridizing an oligonucleotide encoding the desired mutation to a single strand of such starting DNA. After hybridization, a DNA polymerase is used to synthesize an entire second strand, using the hybridized oligonucleotide as a primer, and using the single strand of the starting DNA as a template. Thus, the oligonucleotide encoding the desired mutation is incorporated in the resulting double-stranded DNA.

[107] PCR mutagenesis is also suitable for making amino acid sequence variants of the starting polypeptide, i.e., *H. jecorina* CBH1. See Higuchi, in PCR Protocols, pp.177-183 (Academic Press, 1990); and Vallette et al., Nuc. Acids Res. 17:723-733 (1989). See,

also, for example Cadwell et al., PCR Methods and Applications, Vol 2, 28-33 (1992). Briefly, when small amounts of template DNA are used as starting material in a PCR, primers that differ slightly in sequence from the corresponding region in a template DNA can be used to generate relatively large quantities of a specific DNA fragment that differs from the template sequence only at the positions where the primers differ from the template.

[108] Another method for preparing variants, cassette mutagenesis, is based on the technique described by Wells et al., Gene 34:315-323 (1985). The starting material is the plasmid (or other vector) comprising the starting polypeptide DNA to be mutated. The codon(s) in the starting DNA to be mutated are identified. There must be a unique restriction endonuclease site on each side of the identified mutation site(s). If no such restriction sites exist, they may be generated using the above-described oligonucleotide-mediated mutagenesis method to introduce them at appropriate locations in the starting polypeptide DNA. The plasmid DNA is cut at these sites to linearize it. A double-stranded oligonucleotide encoding the sequence of the DNA between the restriction sites but containing the desired mutation(s) is synthesized using standard procedures, wherein the two strands of the oligonucleotide are synthesized separately and then hybridized together using standard techniques. This double-stranded oligonucleotide is referred to as the cassette. This cassette is designed to have 5' and 3' ends that are compatible with the ends of the linearized plasmid, such that it can be directly ligated to the plasmid. This plasmid now contains the mutated DNA sequence.

[109] Alternatively, or additionally, the desired amino acid sequence encoding a variant CBH I can be determined, and a nucleic acid sequence encoding such amino acid sequence variant can be generated synthetically.

[110] The variant CBH I(s) so prepared may be subjected to further modifications, oftentimes depending on the intended use of the cellulase. Such modifications may involve further alteration of the amino acid sequence, fusion to heterologous polypeptide(s) and/or covalent modifications.

V. *cbh1* Nucleic Acids And CBH1 Polypeptides.

A. Variant *cbh-type* Nucleic acids

[111] The nucleic acid sequence for the wild type *H. jecorina* CBH I is shown in Figure 1. The invention encompasses a nucleic acid molecule encoding the variant cellulases described herein. The nucleic acid may be a DNA molecule.

[112] After the isolation and cloning of the CBH I, other methods known in the art, such as site directed mutagenesis, are used to make the substitutions, additions or deletions

that correspond to substituted amino acids in the expressed CBH I variant. Again, site directed mutagenesis and other methods of incorporating amino acid changes in expressed proteins at the DNA level can be found in Sambrook, *et al.* and Ausubel, *et al.*

[113] After DNA sequences that encode the CBH1 variants have been cloned into DNA constructs, the DNA is used to transform microorganisms. The microorganism to be transformed for the purpose of expressing a variant CBH1 according to the present invention may advantageously comprise a strain derived from *Trichoderma sp.* Thus, a preferred mode for preparing variant CBH1 cellulases according to the present invention comprises transforming a *Trichoderma sp.* host cell with a DNA construct comprising at least a fragment of DNA encoding a portion or all of the variant CBH1. The DNA construct will generally be functionally attached to a promoter. The transformed host cell is then grown under conditions so as to express the desired protein. Subsequently, the desired protein product is purified to substantial homogeneity.

[114] However, it may in fact be that the best expression vehicle for a given DNA encoding a variant CBH1 may differ from *H. jecorina*. Thus, it may be that it will be most advantageous to express a protein in a transformation host that bears phylogenetic similarity to the source organism for the variant CBH1. In an alternative embodiment, *Aspergillus niger* can be used as an expression vehicle. For a description of transformation techniques with *A. niger*, see WO 98/31821, the disclosure of which is incorporated by reference in its entirety.

[115] Accordingly, the present description of a *Trichoderma spp.* expression system is provided for illustrative purposes only and as one option for expressing the variant CBH1 of the invention. One of skill in the art, however, may be inclined to express the DNA encoding variant CBH1 in a different host cell if appropriate and it should be understood that the source of the variant CBH1 should be considered in determining the optimal expression host. Additionally, the skilled worker in the field will be capable of selecting the best expression system for a particular gene through routine techniques utilizing the tools available in the art.

B. Variant CBH1 Polypeptides

[116] The amino acid sequence for the wild type *H. jecorina* CBH I is shown in Figure 1. The variant CBH I polypeptides comprises a substitution or deletion at a position corresponding to one or more of residues S8, Q17, G22, T41, N49, S57, N64, A68, A77, N89, S92, N103, A112, S113, E193, S196, M213, L225, T226, P227, T246, D249, R251, Y252, T255, D257, D259, S278, S279, K286, L288, E295, T296, S297, A299, N301, E325, T332, F338, S342, F352, T356, Y371, T380, Y381, V393, R394, S398, V403, S411,

G430, G440, T445, T462, T484, Q487, and P491 in CBH I from *Hypocrea jecorina.* Furthermore, the variant may further comprises a deletion of residues corresponding to residues 382-393 in CBH I from *Hypocrea jecorina*.

[117] The variant CBH I's of this invention have amino acid sequences that are derived from the amino acid sequence of a precursor CBH I. The amino acid sequence of the CBH I variant differs from the precursor CBH I amino acid sequence by the substitution, deletion or insertion of one or more amino acids of the precursor amino acid sequence. In a preferred embodiment, the precursor CBH I is *Hypocrea jecorina* CBH I. The mature amino acid sequence of *H. jecorina* CBH I is shown in Figure 1. Thus, this invention is directed to CBH I variants which contain amino acid residues at positions which are equivalent to the particular identified residue in *H. jecorina* CBH I. A residue (amino acid) of an CBH I homolog is equivalent to a residue of *Hypocrea jecorina* CBH I if it is either homologous (*i.e.*, corresponding in position in either primary or tertiary structure) or is functionally analogous to a specific residue or portion of that residue in *Hypocrea jecorina* CBH I (*i.e.,* having the same or similar functional capacity to combine, react, or interact chemically or structurally). As used herein, numbering is intended to correspond to that of the mature CBH I amino acid sequence as illustrated in Figure 1. In addition to locations within the precursor CBH I, specific residues in the precursor CBH I corresponding to the amino acid positions that are responsible for instability when the precursor CBH I is under thermal stress are identified herein for substitution or deletion. The amino acid position number (*e.g.*, +51) refers to the number assigned to the mature *Hypocrea jecorina* CBH I sequence presented in Figure 1.

[118] The variant CBH1's of this invention have amino acid sequences that are derived from the amino acid sequence of a precursor *H. jecorina* CBH1. The amino acid sequence of the CBH1 variant differs from the precursor CBH1 amino acid sequence by the substitution, deletion or insertion of one or more amino acids of the precursor amino acid sequence. The mature amino acid sequence of *H. jecorina* CBH1 is shown in Figure 1. Thus, this invention is directed to CBH1 variants which contain amino acid residues at positions which are equivalent to the particular identified residue in *H. jecorina* CBH1. A residue (amino acid) of an CBH1 variant is equivalent to a residue of *Hypocrea jecorina* CBH1 if it is either homologous (*i.e.*, corresponding in position in either primary or tertiary structure) or is functionally analogous to a specific residue or portion of that residue in *Hypocrea jecorina* CBH1 (*i.e.,* having the same or similar functional capacity to combine, react, or interact chemically or structurally). As used herein, numbering is intended to correspond to that of the mature CBH1 amino acid sequence as illustrated in Figure 1. In

addition to locations within the precursor CBH1, specific residues in the precursor CBH1 corresponding to the amino acid positions that are responsible for instability when the precursor CBH1 is under thermal stress are identified herein for substitution or deletion. The amino acid position number (*e.g.*, +51) refers to the number assigned to the mature *Hypocrea jecorina* CBH1 sequence presented in Figure 1.

[119] Alignment of amino acid sequences to determine homology is preferably determined by using a "sequence comparison algorithm." Optimal alignment of sequences for comparison can be conducted, *e.g.,* by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* **2**:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* **48**:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l Acad. Sci. USA* **85**:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), by visual inspection or MOE by Chemical Computing Group, Montreal Canada.

[120] An example of an algorithm that is suitable for determining sequence similarity is the BLAST algorithm, which is described in Altschul, *et al., J. Mol. Biol.* **215**:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<www.ncbi.nlm.nih.gov>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. These initial neighborhood word hits act as starting points to find longer HSPs containing them. The word hits are expanded in both directions along each of the two sequences being compared for as far as the cumulative alignment score can be increased. Extension of the word hits is stopped when: the cumulative alignment score falls off by the quantity X from a maximum achieved value; the cumulative score goes to zero or below; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* **89**:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M'5, N'-4, and a comparison of both strands.

[121] The BLAST algorithm then performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* **90**:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum

probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, an amino acid sequence is considered similar to a protease if the smallest sum probability in a comparison of the test amino acid sequence to a protease amino acid sequence is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

[122] Additional specific strategies for modifying stability of CBH1 cellulases are provided below:

[123] (1) Decreasing the entropy of main-chain unfolding may introduce stability to the enzyme. For example, the introduction of proline residues may significantly stabilize the protein by decreasing the entropy of the unfolding (*see, e.g.,* Watanabe, *et al., Eur. J. Biochem.* **226**:277-283 (1994)). Similarly, glycine residues have no β-carbon, and thus have considerably greater backbone conformational freedom than many other residues. Replacement of glycines, preferably with alanines, may reduce the entropy of unfolding and improve stability (*see, e.g.*, Matthews, *et al.*, *Proc. Natl. Acad. Sci. USA* **84**; 6663-6667 (1987)). Additionally, by shortening external loops it may be possible to improve stability. It has been observed that hyperthermophile produced proteins have shorter external loops than their mesophilic homologues (*see, e.g.,* Russel, *et al., Current Opinions in Biotechnology* **6**:370-374 (1995)). The introduction of disulfide bonds may also be effective to stabilize distinct tertiary structures in relation to each other. Thus, the introduction of cysteines at residues accessible to existing cysteines or the introduction of pairs of cysteines that could form disulfide bonds would alter the stability of a CBH1 variant.

[124] (2) Decreasing internal cavities by increasing side-chain hydrophobicity may alter the stability of an enzyme. Reducing the number and volume of internal cavities increases the stability of enzyme by maximizing hydrophobic interactions and reducing packing defects (*see, e.g.,* Matthews, *Ann. Rev. Biochem.* **62**:139-160 (1993); Burley, *et al., Science* **229**:23-29 (1985); Zuber, *Biophys. Chem.* **29**:171-179 (1988); Kellis, *et al., Nature* **333**:784-786 (1988)). It is known that multimeric proteins from thermophiles often have more hydrophobic sub-unit interfaces with greater surface complementarity than their mesophilic counterparts (Russel, *et al., supra*). This principle is believed to be applicable to domain interfaces of monomeric proteins. Specific substitutions that may improve stability by increasing hydrophobicity include lysine to arginine, serine to alanine and threonine to alanine (Russel, *et al., supra*). Modification by substitution to alanine or proline may increase side-chain size with resultant reduction in cavities, better packing

and increased hydrophobicity. Substitutions to reduce the size of the cavity, increase hydrophobicity and improve the complementarity the interfaces between the domains of CBH1 may improve stability of the enzyme. Specifically, modification of the specific residue at these positions with a different residue selected from any of phenylalanine, tryptophan, tyrosine, leucine and isoleucine may improve performance.

[125] (3) Balancing charge in rigid secondary structure, *i.e.,* α-helices and β-turns may improve stability. For example, neutralizing partial positive charges on a helix N-terminus with negative charge on aspartic acid may improve stability of the structure (*see, e.g.*, Eriksson, *et al., Science* **255**:178-183 (1992)). Similarly, neutralizing partial negative charges on helix C-terminus with positive charge may improve stability. Removing positive charge from interacting with peptide N-terminus in β-turns should be effective in conferring tertiary structure stability. Substitution with a non-positively charged residue could remove an unfavorable positive charge from interacting with an amide nitrogen present in a turn.

[126] (4) Introducing salt bridges and hydrogen bonds to stabilize tertiary structures may be effective. For example, ion pair interactions, *e.g.,* between aspartic acid or glutamic acid and lysine, arginine or histidine, may introduce strong stabilizing effects and may be used to attach different tertiary structure elements with a resultant improvement in thermostability. Additionally, increases in the number of charged residue/non-charged residue hydrogen bonds, and the number of hydrogen-bonds generally, may improve thermostability (*see, e.g.,* Tanner, *et al., Biochemistry* **35**:2597-2609 (1996)). Substitution with aspartic acid, asparagine, glutamic acid or glutamine may introduce a hydrogen bond with a backbone amide. Substitution with arginine may improve a salt bridge and introduce an H-bond into a backbone carbonyl.

[127] (5) Avoiding thermolabile residues in general may increase thermal stability. For example, asparagine and glutamine are susceptible to deamidation and cysteine is susceptible to oxidation at high temperatures. Reducing the number of these residues in sensitive positions may result in improved thermostability (Russel, *et al., supra*). Substitution or deletion by any residue other than glutamine or cysteine may increase stability by avoidance of a thermolabile residue.

[128] (6) Stabilization or destabilization of binding of a ligand that confers modified stability to CBH1 variants. For example, a component of the matrix in which the CBH1 variants of this invention are used may bind to a specific surfactant/thermal sensitivity site of the CBH1 variant. By modifying the site through substitution, binding of the component to the variant may be strengthened or diminished. For example, a non-aromatic residue in

the binding crevice of CBH1 may be substituted with phenylalanine or tyrosine to introduce aromatic side-chain stabilization where interaction of the cellulose substrate may interact favorably with the benzyl rings, increasing the stability of the CBH1 variant.

[129] (7) Increasing the electronegativity of any of the surfactant/ thermal sensitivity ligands may improve stability under surfactant or thermal stress. For example, substitution with phenylalanine or tyrosine may increase the electronegativity of D (aspartate) residues by improving shielding from solvent, thereby improving stability.

C. Anti-CBH Antibodies

[130] The present invention further provides anti-CBH antibodies. The antibodies may be polyclonal, monoclonal, humanized, bispecific or heteroconjugate antibodies.

[131] Methods of preparing polyclonal antibodies are known to the skilled artisan. The immunizing agent may be an CBH polypeptide or a fusion protein thereof. It may be useful to conjugate the antigen to a protein known to be immunogenic in the mammal being immunized. The immunization protocol may be determined by one skilled in the art based on standard protocols or routine experimentation.

[132] Alternatively, the anti-CBH antibodies may be monoclonal antibodies. Monoclonal antibodies may be produced by cells immunized in an animal or using recombinant DNA methods. (See, *e.g.*, Kohler *et al.*, *Nature*, vol. 256, pp. 495-499, August 7, 1975; U.S. Patent No. 4,816,567).

[133] An anti-CBH antibody of the invention may further comprise a humanized or human antibody. The term "humanized antibody" refers to humanized forms of non-human (*e.g.*, murine) antibodies that are chimeric antibodies, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding partial sequences of antibodies) which contain some portion of the sequence derived from non-human antibody. Methods for humanizing non-human antibodies are well known in the art, as further detailed in Jones *et al.*, Nature 321:522-525, 1986; Riechmann *et al.*, Nature, vol. 332, pp. 323-327, 1988; and Verhoeyen *et al.*, Science, vol. 239, pp. 1534-1536, 1988. Methods for producing human antibodies are also known in the art. See, *e.g.*, Jakobovits, A, *et al.*, Annals New York Academy of Sciences, 764:525-535, 1995 and Jakobovits, A, Curr Opin Biotechnol 6(5):561-6, 1995.

VI. Expression Of Recombinant CBH1 Variants

[134] The methods of the invention rely on the use cells to express variant CBH I, with no particular method of CBH I expression required.

[135] The invention provides host cells which have been transduced, transformed or transfected with an expression vector comprising a variant CBH-encoding nucleic acid sequence. The culture conditions, such as temperature, pH and the like, are those previously used for the parental host cell prior to transduction, transformation or transfection and will be apparent to those skilled in the art.

[136] In one approach, a filamentous fungal cell or yeast cell is transfected with an expression vector having a promoter or biologically active promoter fragment or one or more (*e.g.*, a series) of enhancers which functions in the host cell line, operably linked to a DNA segment encoding CBH, such that CBH is expressed in the cell line.

A. Nucleic Acid Constructs/Expression Vectors.

[137] Natural or synthetic polynucleotide fragments encoding CBH I ("CBH I-encoding nucleic acid sequences") may be incorporated into heterologous nucleic acid constructs or vectors, capable of introduction into, and replication in, a filamentous fungal or yeast cell. The vectors and methods disclosed herein are suitable for use in host cells for the expression of CBH I. Any vector may be used as long as it is replicable and viable in the cells into which it is introduced. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. Cloning and expression vectors are also described in Sambrook *et al.,* 1989, Ausubel FM *et al.,* 1989, and Strathern *et al.*, The Molecular Biology of the Yeast *Saccharomyces*, 1981, each of which is expressly incorporated by reference herein. Appropriate expression vectors for fungi are described in van den Hondel, C.A.M.J.J. et al. (1991) In: Bennett, J.W. and Lasure, L.L. (eds.) More Gene Manipulations in Fungi. Academic Press, pp. 396-428. The appropriate DNA sequence may be inserted into a plasmid or vector (collectively referred to herein as "vectors") by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by standard procedures. Such procedures and related sub-cloning procedures are deemed to be within the scope of knowledge of those skilled in the art.

[138] Recombinant filamentous fungi comprising the coding sequence for variant CBH I may be produced by introducing a heterologous nucleic acid construct comprising the variant CBH I coding sequence into the cells of a selected strain of the filamentous fungi.

[139] Once the desired form of a variant *cbh* nucleic acid sequence is obtained, it may be modified in a variety of ways. Where the sequence involves non-coding flanking regions, the flanking regions may be subjected to resection, mutagenesis, etc. Thus, transitions, transversions, deletions, and insertions may be performed on the naturally occurring sequence.

[140] A selected variant *cbh* coding sequence may be inserted into a suitable vector according to well-known recombinant techniques and used to transform filamentous fungi capable of CBH I expression. Due to the inherent degeneracy of the genetic code, other nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used to clone and express variant CBH I. Therefore it is appreciated that such substitutions in the coding region fall within the sequence variants covered by the present invention. Any and all of these sequence variants can be utilized in the same way as described herein for a parent CBH I-encoding nucleic acid sequence.

[141] The present invention also includes recombinant nucleic acid constructs comprising one or more of the variant CBH I-encoding nucleic acid sequences as described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation.

[142] Heterologous nucleic acid constructs may include the coding sequence for variant *cbh*: (i) in isolation; (ii) in combination with additional coding sequences; such as fusion protein or signal peptide coding sequences, where the *cbh* coding sequence is the dominant coding sequence; (iii) in combination with non-coding sequences, such as introns and control elements, such as promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in a suitable host; and/or (iv) in a vector or host environment in which the *cbh* coding sequence is a heterologous gene.

[143] In one aspect of the present invention, a heterologous nucleic acid construct is employed to transfer a variant CBH I-encoding nucleic acid sequence into a cell *in vitro*, with established filamentous fungal and yeast lines preferred. For long-term, production of variant CBH I, stable expression is preferred. It follows that any method effective to generate stable transformants may be used in practicing the invention.

[144] Appropriate vectors are typically equipped with a selectable marker-encoding nucleic acid sequence, insertion sites, and suitable control elements, such as promoter and termination sequences. The vector may comprise regulatory sequences, including, for example, non-coding sequences, such as introns and control elements, *i.e.*, promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in host cells (and/or in a vector or host cell environment in which a modified soluble protein antigen coding sequence is not normally expressed), operably linked to the coding sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, many of which are commercially available and/or are described in Sambrook, *et al.*, (*supra*).

[145] Exemplary promoters include both constitutive promoters and inducible promoters, examples of which include a CMV promoter, an SV40 early promoter, an RSV promoter, an EF-1α promoter, a promoter containing the tet responsive element (TRE) in the tet-on or tet-off system as described (ClonTech and BASF), the beta actin promoter and the metallothionine promoter that can upregulated by addition of certain metal salts. A promoter sequence is a DNA sequence which is recognized by the particular filamentous fungus for expression purposes. It is operably linked to DNA sequence encoding a variant CBH I polypeptide. Such linkage comprises positioning of the promoter with respect to the initiation codon of the DNA sequence encoding the variant CBH I polypeptide in the disclosed expression vectors. The promoter sequence contains transcription and translation control sequence which mediate the expression of the variant CBH I polypeptide. Examples include the promoters from the *Aspergillus niger*, *A awamori* or *A. oryzae* glucoamylase, alpha-amylase, or alpha-glucosidase encoding genes; the *A. nidulans gpdA* or *trpC* Genes; the *Neurospora crassa cbh1* or *trp1* genes; the *A. niger* or *Rhizomucor miehei* aspartic proteinase encoding genes; the *H. jecorina* (*T. reesei) cbh1*, *cbh2*, *egl1*, *egl2*, or other cellulase encoding genes.

[146] The choice of the proper selectable marker will depend on the host cell, and appropriate markers for different hosts are well known in the art. Typical selectable marker genes include argB from *A. nidulans* or *T. reesei*, amdS from *A. nidulans*, pyr4 from *Neurospora crassa* or *T. reesei*, pyrG from *Aspergillus niger or A. nidulans*. Additional exemplary selectable markers include, but are not limited to trpc, trp1, oliC31, niaD or leu2, which are included in heterologous nucleic acid constructs used to transform a mutant strain such as trp-, pyr-, leu- and the like.

[147] Such selectable markers confer to transformants the ability to utilize a metabolite that is usually not metabolized by the filamentous fungi. For example, the amdS gene from *H. jecorina* which encodes the enzyme acetamidase that allows transformant cells to grow on acetamide as a nitrogen source. The selectable marker (e.g. pyrG) may restore the ability of an auxotrophic mutant strain to grow on a selective minimal medium or the selectable marker (e.g. olic31) may confer to transformants the ability to grow in the presence of an inhibitory drug or antibiotic.

[148] The selectable marker coding sequence is cloned into any suitable plasmid using methods generally employed in the art. Exemplary plasmids include pUC18, pBR322, pRAX and pUC100. The pRAX plasmid contains AMA1 sequences from A. nidulans, which make it possible to replicate in A. niger.

[149] The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Sambrook *et al.*, 1989; Freshney, Animal Cell Culture, 1987; Ausubel, *et al.*, 1993; and Coligan *et al.*, Current Protocols in Immunology, 1991.

B. Host Cells and Culture Conditions For CBH1 Production

(i) Filamentous Fungi

[150] Thus, the present invention provides filamentous fungi comprising cells which have been modified, selected and cultured in a manner effective to result in variant CBH I production or expression relative to the corresponding non-transformed parental fungi.

[151] Examples of species of parental filamentous fungi that may be treated and/or modified for variant CBH I expression include, but are not limited to *Trichoderma, e.g., Trichoderma reesei, Trichoderma longibrachiatum , Trichoderma viride, Trichoderma koningii; Penicillium sp., Humicola sp.*, including *Humicola insolens; Aspergillus sp., Chrysosporium sp., Fusarium sp., Hypocrea* sp., and *Emericella* sp.

[152] CBH I expressing cells are cultured under conditions typically employed to culture the parental fungal line. Generally, cells are cultured in a standard medium containing physiological salts and nutrients, such as described in Pourquie, J. et al., Biochemistry and Genetics of Cellulose Degradation, eds. Aubert, J. P. et al., Academic Press, pp. 71-86, 1988 and Ilmen, M. et al., Appl. Environ. Microbiol. 63:1298-1306, 1997. Culture conditions are also standard, *e.g.*, cultures are incubated at 28°C in shaker cultures or fermenters until desired levels of CBH I expression are achieved.

[153] Preferred culture conditions for a given filamentous fungus may be found in the scientific literature and/or from the source of the fungi such as the American Type Culture Collection (ATCC; "http://www.atcc.org/"). After fungal growth has been established, the cells are exposed to conditions effective to cause or permit the expression of variant CBH I.

[154] In cases where a CBH I coding sequence is under the control of an inducible promoter, the inducing agent, *e.g.*, a sugar, metal salt or antibiotics, is added to the medium at a concentration effective to induce CBH I expression.

[155] In one embodiment, the strain comprises *Aspergillus niger,* which is a useful strain for obtaining overexpressed protein. For example A. niger var awamori dgr246 is known to secrete elevated amounts of secreted cellulases (Goedegebuur et al, Curr. Genet (2002) 41: 89-98). Other strains of Aspergillus niger var awamori such as GCDAP3,

GCDAP4 and GAP3-4 are known Ward et al (Ward, M, Wilson, L.J. and Kodama, K.H., 1993, Appl. Microbiol. Biotechnol. 39:738-743).

[156] In another embodiment, the strain comprises *Trichoderma reesei,* which is a useful strain for obtaining overexpressed protein. For example, RL-P37, described by Sheir-Neiss, *et al., Appl. Microbiol. Biotechnol.* 20:46-53 (1984) is known to secrete elevated amounts of cellulase enzymes. Functional equivalents of RL-P37 include *Trichoderma reesei* strain RUT-C30 (ATCC No. 56765) and strain QM9414 (ATCC No. 26921). It is contemplated that these strains would also be useful in overexpressing variant CBH1.

[157] Where it is desired to obtain the variant CBH I in the absence of potentially detrimental native cellulolytic activity, it is useful to obtain a *Trichoderma* host cell strain which has had one or more cellulase genes deleted prior to introduction of a DNA construct or plasmid containing the DNA fragment encoding the variant CBH I. Such strains may be prepared by the method disclosed in U.S. Patent No. 5,246,853 and WO 92/06209, which disclosures are hereby incorporated by reference. By expressing a variant CBH I cellulase in a host microorganism that is missing one or more cellulase genes, the identification and subsequent purification procedures are simplified. Any gene from *Trichoderma sp.* which has been cloned can be deleted, for example, the *cbh1*, *cbh2*, *egl1*, and *egl2* genes as well as those encoding EG III and/or EGV protein (*see e.g.*, U.S. Patent No. 5,475,101 and WO 94/28117, respectively).

[158] Gene deletion may be accomplished by inserting a form of the desired gene to be deleted or disrupted into a plasmid by methods known in the art. The deletion plasmid is then cut at an appropriate restriction enzyme site(s), internal to the desired gene coding region, and the gene coding sequence or part thereof replaced with a selectable marker. Flanking DNA sequences from the locus of the gene to be deleted or disrupted, preferably between about 0.5 to 2.0 kb, remain on either side of the selectable marker gene. An appropriate deletion plasmid will generally have unique restriction enzyme sites present therein to enable the fragment containing the deleted gene, including flanking DNA sequences, and the selectable marker gene to be removed as a single linear piece.

[159] A selectable marker must be chosen so as to enable detection of the transformed microorganism. Any selectable marker gene that is expressed in the selected microorganism will be suitable. For example, with *Aspergillus sp.*, the selectable marker is chosen so that the presence of the selectable marker in the transformants will not significantly affect the properties thereof. Such a selectable marker may be a gene that encodes an assayable product. For example, a functional copy of a *Aspergillus sp.* gene

may be used which if lacking in the host strain results in the host strain displaying an auxotrophic phenotype. Similarly, selectable markers exist for *Trichoderma sp.*

[160] In one embodiment, a *pyrG*⁻ derivative strain of *Aspergillus sp.* is transformed with a functional *pyrG* gene, which thus provides a selectable marker for transformation. A *pyrG*⁻ derivative strain may be obtained by selection of *Aspergillus sp.* strains that are resistant to fluoroorotic acid (FOA). The *pyrG* gene encodes orotidine-5'-monophosphate decarboxylase, an enzyme required for the biosynthesis of uridine. Strains with an intact *pyrG* gene grow in a medium lacking uridine but are sensitive to fluoroorotic acid. It is possible to select *pyrG*⁻ derivative strains that lack a functional orotidine monophosphate decarboxylase enzyme and require uridine for growth by selecting for FOA resistance. Using the FOA selection technique it is also possible to obtain uridine-requiring strains which lack a functional orotate pyrophosphoribosyl transferase. It is possible to transform these cells with a functional copy of the gene encoding this enzyme (Berges & Barreau, *Curr. Genet.* 19:359-365 (1991), and van Hartingsveldte et al., (1986) Development of a homologous transformation system for *Aspergillus niger* based on the *pyrG* gene. Mol. Gen. Genet. 206:71-75). Selection of derivative strains is easily performed using the FOA resistance technique referred to above, and thus, the *pyrG* gene is preferably employed as a selectable marker.

[161] In a second embodiment, a *pyr4*⁻ derivative strain of *Hyprocrea sp. (Hyprocrea sp. (Trichoderma sp.))* is transformed with a functional *pyr4* gene, which thus provides a selectable marker for transformation. A *pyr4*⁻ derivative strain may be obtained by selection of *Hyprocrea sp. (Trichoderma sp.)* strains that are resistant to fluoroorotic acid (FOA). The *pyr4* gene encodes orotidine-5'-monophosphate decarboxylase, an enzyme required for the biosynthesis of uridine. Strains with an intact *pyr4* gene grow in a medium lacking uridine but are sensitive to fluoroorotic acid. It is possible to select *pyr4*⁻ derivative strains that lack a functional orotidine monophosphate decarboxylase enzyme and require uridine for growth by selecting for FOA resistance. Using the FOA selection technique it is also possible to obtain uridine-requiring strains which lack a functional orotate pyrophosphoribosyl transferase. It is possible to transform these cells with a functional copy of the gene encoding this enzyme (Berges & Barreau, *Curr. Genet.* **19**:359-365 (1991)). Selection of derivative strains is easily performed using the FOA resistance technique referred to above, and thus, the *pyr4* gene is preferably employed as a selectable marker.

[162] To transform *pyrG*⁻ *Aspergillus sp.* or *pyr4*⁻ *Hyprocrea sp. (Trichoderma sp.)* so as to be lacking in the ability to express one or more cellulase genes, a single DNA fragment

comprising a disrupted or deleted cellulase gene is then isolated from the deletion plasmid and used to transform an appropriate *pyr⁻ Aspergillus* or *pyr⁻ Trichoderma* host. Transformants are then identified and selected based on their ability to express the *pyrG* or *pyr4,* respecitively, gene product and thus compliment the uridine auxotrophy of the host strain. Southern blot analysis is then carried out on the resultant transformants to identify and confirm a double crossover integration event that replaces part or all of the coding region of the genomic copy of the gene to be deleted with the appropriate *pyr* selectable markers.

[163] Although the specific plasmid vectors described above relate to preparation of *pyr⁻* transformants, the present invention is not limited to these vectors. Various genes can be deleted and replaced in the *Aspergillus sp. or Hyprocrea sp. (Trichoderma sp.)* strain using the above techniques. In addition, any available selectable markers can be used, as discussed above. In fact, any host, e.g., *Aspergillus sp. or Hyprocrea sp.,* gene that has been cloned, and thus identified, can be deleted from the genome using the above-described strategy.

[164] As stated above, the host strains used may be derivatives of *Hyprocrea sp. (Trichoderma sp.)* that lack or have a nonfunctional gene or genes corresponding to the selectable marker chosen. For example, if the selectable marker of *pyrG* is chosen for *Aspergillus sp.,* then a specific *pyrG⁻* derivative strain is used as a recipient in the transformation procedure. Also, for example, if the selectable marker of *pyr4* is chosen for a *Hyprocrea sp.*, then a specific *pyr4⁻* derivative strain is used as a recipient in the transformation procedure. Similarly, selectable markers comprising *Hyprocrea sp. (Trichoderma sp.)* genes equivalent to the *Aspergillus nidulans* genes *amdS, argB, trpC, niaD* may be used. The corresponding recipient strain must therefore be a derivative strain such as *argB⁻, trpC⁻, niaD⁻*, respectively.

[165] DNA encoding the CBH I variant is then prepared for insertion into an appropriate microorganism. According to the present invention, DNA encoding a CBH I variant comprises the DNA necessary to encode for a protein that has functional cellulolytic activity. The DNA fragment encoding the CBH I variant may be functionally attached to a fungal promoter sequence, for example, the promoter of the *glaA* gene in *Aspergillus* or the promoter of the *cbh1* or *egl1* genes in *Trichoderma*.

[166] It is also contemplated that more than one copy of DNA encoding a CBH I variant may be recombined into the strain to facilitate overexpression. The DNA encoding the CBH I variant may be prepared by the construction of an expression vector carrying the DNA encoding the variant. The expression vector carrying the inserted DNA fragment

encoding the CBH I variant may be any vector which is capable of replicating autonomously in a given host organism or of integrating into the DNA of the host, typically a plasmid. In preferred embodiments two types of expression vectors for obtaining expression of genes are contemplated. The first contains DNA sequences in which the promoter, gene-coding region, and terminator sequence all originate from the gene to be expressed. Gene truncation may be obtained where desired by deleting undesired DNA sequences (*e.g.*, coding for unwanted domains) to leave the domain to be expressed under control of its own transcriptional and translational regulatory sequences. A selectable marker may also be contained on the vector allowing the selection for integration into the host of multiple copies of the novel gene sequences.

[167] The second type of expression vector is preassembled and contains sequences required for high-level transcription and a selectable marker. It is contemplated that the coding region for a gene or part thereof can be inserted into this general-purpose expression vector such that it is under the transcriptional control of the expression cassettes promoter and terminator sequences.

[168] For example, in *Aspergillus,* pRAX is such a general-purpose expression vector. Genes or part thereof can be inserted downstream of the strong *glaA* promoter.

[169] For example, in *Hypocrea,* pTEX is such a general-purpose expression vector. Genes or part thereof can be inserted downstream of the strong *cbh*1 promoter.

[170] In the vector, the DNA sequence encoding the CBH I variant of the present invention should be operably linked to transcriptional and translational sequences, *i.e.,* a suitable promoter sequence and signal sequence in reading frame to the structural gene. The promoter may be any DNA sequence that shows transcriptional activity in the host cell and may be derived from genes encoding proteins either homologous or heterologous to the host cell. An optional signal peptide provides for extracellular production of the CBH I variant. The DNA encoding the signal sequence is preferably that which is naturally associated with the gene to be expressed, however the signal sequence from any suitable source, for example an exo-cellobiohydrolase or endoglucanase from *Trichoderma*, is contemplated in the present invention.

[171] The procedures used to ligate the DNA sequences coding for the variant CBH I of the present invention with the promoter, and insertion into suitable vectors are well known in the art.

[172] The DNA vector or construct described above may be introduced in the host cell in accordance with known techniques such as transformation, transfection, microinjection, microporation, biolistic bombardment and the like.

[173] In the preferred transformation technique, it must be taken into account that the permeability of the cell wall to DNA in *Hyprocrea sp. (Trichoderma sp.)* is very low. Accordingly, uptake of the desired DNA sequence, gene or gene fragment is at best minimal. There are a number of methods to increase the permeability of the *Hyprocrea sp. (Trichoderma sp.)* cell wall in the derivative strain (*i.e.*, lacking a functional gene corresponding to the used selectable marker) prior to the transformation process.

[174] The preferred method in the present invention to prepare *Aspergillus sp.* or *Hyprocrea sp. (Trichoderma sp.)* for transformation involves the preparation of protoplasts from fungal mycelium. See Campbell *et al.* Improved transformation efficiency of *A.niger* using homologous *nia*D gene for nitrate reductase. Curr. Genet. 16:53-56; 1989. The mycelium can be obtained from germinated vegetative spores. The mycelium is treated with an enzyme that digests the cell wall resulting in protoplasts. The protoplasts are then protected by the presence of an osmotic stabilizer in the suspending medium. These stabilizers include sorbitol, mannitol, potassium chloride, magnesium sulfate and the like. Usually the concentration of these stabilizers varies between 0.8 M and 1.2 M. It is preferable to use about a 1.2 M solution of sorbitol in the suspension medium.

[175] Uptake of the DNA into the host strain, (*Aspergillus sp.* or *Hyprocrea sp. (Trichoderma sp.),* is dependent upon the calcium ion concentration. Generally between about 10 mM $CaCl_2$ and 50 mM $CaCl_2$ is used in an uptake solution. Besides the need for the calcium ion in the uptake solution, other items generally included are a buffering system such as TE buffer (10 Mm Tris, pH 7.4; 1 mM EDTA) or 10 mM MOPS, pH 6.0 buffer (morpholinepropanesulfonic acid) and polyethylene glycol (PEG). It is believed that the polyethylene glycol acts to fuse the cell membranes thus permitting the contents of the medium to be delivered into the cytoplasm of the host cell, by way of example either *Aspergillus sp.* or *Hyprocrea sp.* strain, and the plasmid DNA is transferred to the nucleus. This fusion frequently leaves multiple copies of the plasmid DNA tenderly integrated into the host chromosome.

[176] Usually a suspension containing the *Aspergillus sp.* protoplasts or cells that have been subjected to a permeability treatment at a density of 10^5 to 10^6/mL, preferably 2 x 10^5/mL are used in transformation. Similarly, a suspension containing the *Hyprocrea sp. (Trichoderma sp.)* protoplasts or cells that have been subjected to a permeability treatment at a density of 10^8 to 10^9/mL, preferably 2 x 10^8/mL are used in transformation. A volume of 100 µL of these protoplasts or cells in an appropriate solution (*e.g.*, 1.2 M sorbitol; 50 mM $CaCl_2$) are mixed with the desired DNA. Generally a high concentration of PEG is added to the uptake solution. From 0.1 to 1 volume of 25% PEG 4000 can be

added to the protoplast suspension. However, it is preferable to add about 0.25 volumes to the protoplast suspension. Additives such as dimethyl sulfoxide, heparin, spermidine, potassium chloride and the like may also be added to the uptake solution and aid in transformation.

[177] Generally, the mixture is then incubated at approximately 0°C for a period of between 10 to 30 minutes. Additional PEG is then added to the mixture to further enhance the uptake of the desired gene or DNA sequence. The 25% PEG 4000 is generally added in volumes of 5 to 15 times the volume of the transformation mixture; however, greater and lesser volumes may be suitable. The 25% PEG 4000 is preferably about 10 times the volume of the transformation mixture. After the PEG is added, the transformation mixture is then incubated either at room temperature or on ice before the addition of a sorbitol and $CaCl_2$ solution. The protoplast suspension is then further added to molten aliquots of a growth medium. This growth medium permits the growth of transformants only. Any growth medium can be used in the present invention that is suitable to grow the desired transformants. However, if *Pyr*$^+$ transformants are being selected it is preferable to use a growth medium that contains no uridine. The subsequent colonies are transferred and purified on a growth medium depleted of uridine.

[178] At this stage, stable transformants may be distinguished from unstable transformants by their faster growth rate and the formation of circular colonies with a smooth, rather than ragged outline on solid culture medium lacking uridine. Additionally, in some cases a further test of stability may made by growing the transformants on solid non-selective medium (*i.e.* containing uridine), harvesting spores from this culture medium and determining the percentage of these spores which will subsequently germinate and grow on selective medium lacking uridine.

[179] In a particular embodiment of the above method, the CBH I variant(s) are recovered in active form from the host cell after growth in liquid media either as a result of the appropriate post translational processing of the CBH I variant.

(ii) Yeast

[180] The present invention also contemplates the use of yeast as a host cell for CBH I production. Several other genes encoding hydrolytic enzymes have been expressed in various strains of the yeast *S. cerevisiae*. These include sequences encoding for two endoglucanases (Penttila *et al.*, Yeast vol. 3, pp 175-185, 1987), two cellobiohydrolases (Penttila *et al.*, Gene, 63: 103-112, 1988) and one beta-glucosidase from *Trichoderma reesei* (Cummings and Fowler, Curr. Genet. 29:227-233, 1996), a xylanase from *Aureobasidlium pullulans* (Li and Ljungdahl, Appl. Environ. Microbiol. 62, no. 1, pp. 209-

213, 1996), an alpha-amylase from wheat (Rothstein *et al.*, Gene 55:353-356, 1987), etc. In addition, a cellulase gene cassette encoding the *Butyrivibrio fibrisolvens* endo- [beta] -1,4-glucanase (END1), *Phanerochaete chrysosporium* cellobiohydrolase (CBH1), the *Ruminococcus flavefaciens* cellodextrinase (CEL1) and the *Endomyces fibrilizer* cellobiase (Bgl1) was successfully expressed in a laboratory strain of *S. cerevisiae* (Van Rensburg *et al.*, Yeast, vol. 14, pp. 67-76, 1998).

C. Introduction of an CBH I-Encoding Nucleic Acid Sequence into Host Cells.

[181] The invention further provides cells and cell compositions which have been genetically modified to comprise an exogenously provided variant CBH I -encoding nucleic acid sequence. A parental cell or cell line may be genetically modified (*i.e.*, transduced, transformed or transfected) with a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc, as further described above.

[182] The methods of transformation of the present invention may result in the stable integration of all or part of the transformation vector into the genome of the filamentous fungus. However, transformation resulting in the maintenance of a self-replicating extra-chromosomal transformation vector is also contemplated.

[183] Many standard transfection methods can be used to produce *Trichoderma reesei* cell lines that express large quantities of the heterologus protein. Some of the published methods for the introduction of DNA constructs into cellulase-producing strains of Trichoderma include Lorito, Hayes, DiPietro and Harman, 1993, Curr. Genet. 24: 349-356; Goldman, VanMontagu and Herrera-Estrella, 1990, Curr. Genet. 17:169-174; Penttila, Nevalainen, Ratto, Salminen and Knowles, 1987, Gene 6: 155-164, for *Aspergillus* Yelton, Hamer and Timberlake, 1984, Proc. Natl. Acad. Sci. USA 81: 1470-1474, for Fusarium Bajar, Podila and Kolattukudy, 1991, Proc. Natl. Acad. Sci. USA 88: 8202-8212, for Streptomyces Hopwood et al., 1985, The John Innes Foundation, Norwich, UK and for Bacillus Brigidi, DeRossi, Bertarini, Riccardi and Matteuzzi, 1990, FEMS Microbiol. Lett. 55: 135-138).

[184] Other methods for introducing a heterologous nucleic acid construct (expression vector) into filamentous fungi (*e.g.*, *H. jecorina*) include, but are not limited to the use of a particle or gene gun, permeabilization of filamentous fungi cells walls prior to the transformation process (*e.g.*, by use of high concentrations of alkali, *e.g.*, 0.05 M to 0.4 M $CaC1_2$ or lithium acetate), protoplast fusion or agrobacterium mediated transformation. An exemplary method for transformation of filamentous fungi by treatment of protoplasts or

spheroplasts with polyethylene glycol and $CaCl_2$ is described in Campbell, E.I. et al., Curr. Genet. 16:53-56, 1989 and Penttila, M. et al., Gene, 63:11-22, 1988.

[185] Any of the well-known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, biolistics, liposomes, microinjection, plasma vectors, viral vectors and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (*see, e.g.*, Sambrook *et al., supra*). Also of use is the Agrobacterium-mediated transfection method described in U.S. Patent No. 6,255,115. It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the heterologous gene.

[186] In addition, heterologous nucleic acid constructs comprising a variant CBH I-encoding nucleic acid sequence can be transcribed *in vitro*, and the resulting RNA introduced into the host cell by well-known methods, *e.g.*, by injection.

[187] The invention further includes novel and useful transformants of filamentous fungi such as *H. jecorina* and *A. niger* for use in producing fungal cellulase compositions. The invention includes transformants of filamentous fungi especially fungi comprising the variant CBH I coding sequence, or deletion of the endogenous *cbh* coding sequence.

[188] Following introduction of a heterologous nucleic acid construct comprising the coding sequence for a variant *cbh 1*, the genetically modified cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying expression of a variant CBH I-encoding nucleic acid sequence. The culture conditions, such as temperature, pH and the like, are those previously used for the host cell selected for expression, and will be apparent to those skilled in the art.

[189] The progeny of cells into which such heterologous nucleic acid constructs have been introduced are generally considered to comprise the variant CBH I-encoding nucleic acid sequence found in the heterologous nucleic acid construct.

[190] The invention further includes novel and useful transformants of filamentous fungi such as *H. jecorina* for use in producing fungal cellulase compositions. The invention includes transformants of filamentous fungi especially fungi comprising the variant *cbh 1* coding sequence, or deletion of the endogenous *cbh* coding sequence.

[191] Stable transformants of filamentous fungi can generally be distinguished from unstable transformants by their faster growth rate and the formation of circular colonies with a smooth rather than ragged outline on solid culture medium. Additionally, in some

cases, a further test of stability can be made by growing the transformants on solid non-selective medium, harvesting the spores from this culture medium and determining the percentage of these spores which will subsequently germinate and grow on selective medium.

VII. Analysis For CBH1 Nucleic Acid Coding Sequences and/or Protein Expression.

[192] In order to evaluate the expression of a variant CBH I by a cell line that has been transformed with a variant CBH I-encoding nucleic acid construct, assays can be carried out at the protein level, the RNA level or by use of functional bioassays particular to cellobiohydrolase activity and/or production.

[193] In one exemplary application of the variant *cbh 1* nucleic acid and protein sequences described herein, a genetically modified strain of filamentous fungi, *e.g.*, *Trichoderma reesei*, is engineered to produce an increased amount of CBH I. Such genetically modified filamentous fungi would be useful to produce a cellulase product with greater increased cellulolytic capacity. In one approach, this is accomplished by introducing the coding sequence for *cbh 1* into a suitable host, *e.g.*, a filamentous fungi such as *Aspergillus niger*.

[194] Accordingly, the invention includes methods for expressing variant CBH I in a filamentous fungus or other suitable host by introducing an expression vector containing the DNA sequence encoding variant CBH I into cells of the filamentous fungus or other suitable host.

[195] In another aspect, the invention includes methods for modifying the expression of CBH I in a filamentous fungus or other suitable host. Such modification includes a decrease or elimination in expression of the endogenous CBH.

[196] In general, assays employed to analyze the expression of variant CBH I include, Northern blotting, dot blotting (DNA or RNA analysis), RT-PCR (reverse transcriptase polymerase chain reaction), or *in situ* hybridization, using an appropriately labeled probe (based on the nucleic acid coding sequence) and conventional Southern blotting and autoradiography.

[197] In addition, the production and/or expression of variant CBH I may be measured in a sample directly, for example, by assays for cellobiohydrolase activity, expression and/or production. Such assays are described, for example, in Becker et al., Biochem J. (2001) 356:19-30 and Mitsuishi et al., FEBS (1990) 275:135-138, each of which is expressly incorporated by reference herein. The ability of CBH I to hydrolyze isolated soluble and insoluble substrates can be measured using assays described in Srisodsuk et al., J.

Biotech. (1997) 57:49-57 and Nidetzky and Claeyssens Biotech. Bioeng. (1994) 44:961-966. Substrates useful for assaying cellobiohydrolase, endoglucanase or β-glucosidase activities include crystalline cellulose, filter paper, phosphoric acid swollen cellulose, cellooligosaccharides, methylumbelliferyl lactoside, methylumbelliferyl cellobioside, orthonitrophenyl lactoside, paranitrophenyl lactoside, orthonitrophenyl cellobioside, paranitrophenyl cellobioside.

[198] In addition, protein expression, may be evaluated by immunological methods, such as immunohistochemical staining of cells, tissue sections or immunoassay of tissue culture medium, *e.g.*, by Western blot or ELISA. Such immunoassays can be used to qualitatively and quantitatively evaluate expression of a CBH I variant. The details of such methods are known to those of skill in the art and many reagents for practicing such methods are commercially available.

[199] A purified form of a variant CBH I may be used to produce either monoclonal or polyclonal antibodies specific to the expressed protein for use in various immunoassays. (See, *e.g.*, Hu *et al.*, Mol Cell Biol. vol.11, no. 11, pp. 5792-5799, 1991). Exemplary assays include ELISA, competitive immunoassays, radioimmunoassays, Western blot, indirect immunofluorescent assays and the like. In general, commercially available antibodies and/or kits may be used for the quantitative immunoassay of the expression level of cellobiohydrolase proteins.

VIII. Isolation And Purification Of Recombinant CBH1 Protein.

[200] In general, a variant CBH I protein produced in cell culture is secreted into the medium and may be purified or isolated, *e.g.*, by removing unwanted components from the cell culture medium. However, in some cases, a variant CBH I protein may be produced in a cellular form necessitating recovery from a cell lysate. In such cases the variant CBH I protein is purified from the cells in which it was produced using techniques routinely employed by those of skill in the art. Examples include, but are not limited to, affinity chromatography (Tilbeurgh *et al.*, FEBS Lett. 16:215, 1984), ion-exchange chromatographic methods (Goyal *et al.*, Bioresource Technol. 36:37-50, 1991; Fliess *et al.*, Eur. J. Appl. Microbiol. Biotechnol. 17:314-318, 1983; Bhikhabhai *et al.*, J. Appl. Biochem. 6:336-345, 1984; Ellouz *et al.*, J. Chromatography 396:307-317, 1987), including ion-exchange using materials with high resolution power (Medve *et al.*, J. Chromatography A 808:153-165, 1998), hydrophobic interaction chromatography (Tomaz and Queiroz, J. Chromatography A 865:123-128, 1999), and two-phase partitioning (Brumbauer, *et al.*, Bioseparation 7:287-295, 1999).

[201] Typically, the variant CBH I protein is fractionated to segregate proteins having selected properties, such as binding affinity to particular binding agents, *e.g.*, antibodies or receptors; or which have a selected molecular weight range, or range of isoelectric points.

[202] Once expression of a given variant CBH I protein is achieved, the CBH I protein thereby produced is purified from the cells or cell culture. Exemplary procedures suitable for such purification include the following: antibody-affinity column chromatography, ion exchange chromatography; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; and gel filtration using, *e.g.*, Sephadex G-75. Various methods of protein purification may be employed and such methods are known in the art and described *e.g.* in Deutscher, Methods in Enzymology, vol. 182, no. 57, pp. 779, 1990; Scopes, Methods Enzymol. 90: 479-91, 1982. The purification step(s) selected will depend, *e.g.*, on the nature of the production process used and the particular protein produced.

IX. Utility of *cbh1* and CBH1

[203] **It can be appreciated that** the variant *cbh* nucleic acids, the variant CBH I protein and compositions comprising variant CBH I protein activity find utility in a wide variety applications, some of which are described below.

[204] New and improved cellulase compositions that comprise varying amounts BG-type, EG-type and variant CBH-type cellulases find utility in detergent compositions that exhibit enhanced cleaning ability, function as a softening agent and/or improve the feel of cotton fabrics (*e.g.*, "stone washing" or "biopolishing"), in compositions for degrading wood pulp into sugars (*e.g.*, for bio-ethanol production), and/or in feed compositions. The isolation and characterization of cellulase of each type provides the ability to control the aspects of such compositions.

[205] Variant (or mutant) CBHs with increased thermostability find uses in all of the above areas due to their ability to retain activity at elevated temperatures.

[206] Variant (or mutant) CBHs with decreased thermostability find uses, for example, in areas where the enzyme activity is required to be neutralized at lower temperatures so that other enzymes that may be present are left unaffected. In addition, the enzymes may find utility in the limited conversion of cellulosics, for example, in controlling the degree of crystallinity or of cellulosic chain-length. After reaching the desired extent of conversion the saccharifying temperature can be raised above the survival temperature of the de-stabilized CBH I. As the CBH I activity is essential for hydrolysis of crystalline cellulose, conversion of crystalline cellulose will cease at the elevated temperature.

[207] Variant (or mutant) CBHs with increased reversibility, i.e., enhanced refolding and retention of activity, also find use in similar areas. Depending upon the conditions of thermal inactivation, reversible denaturation can compete with, or dominate over, the irreversible process. Variants with increased reversibility would, under these conditions, exhibit increased resistance to thermal inactivation. Increased reversibility would also be of potential benefit in any process in which an inactivation event was followed by a treatment under non-inactivating conditions. For instance, in a Hybrid Hydrolysis and Fermentation (HHF) process for biomass conversion to ethanol, the biomass would first be incompletely saccharified by cellulases at elevated temperature (say 50°C or higher), then the temperature would be dropped (to 30°C, for instance) to allow a fermentative organism to be introduced to convert the sugars to ethanol. If, upon decrease of process temperature, thermally inactivated cellulase reversibly re-folded and recovered activity then saccharification could continue to higher levels of conversion during the low temperature fermentation process.

[208] In one approach, the cellulase of the invention finds utility in detergent compositions or in the treatment of fabrics to improve the feel and appearance.

[209] Since the rate of hydrolysis of cellulosic products may be increased by using a transformant having at least one additional copy of the *cbh* gene inserted into the genome, products that contain cellulose or heteroglycans can be degraded at a faster rate and to a greater extent. Products made from cellulose such as paper, cotton, cellulosic diapers and the like can be degraded more efficiently in a landfill. Thus, the fermentation product obtainable from the transformants or the transformants alone may be used in compositions to help degrade by liquefaction a variety of cellulose products that add to the overcrowded landfills.

[210] Separate saccharification and fermentation is a process whereby cellulose present in biomass, e.g., corn stover, is converted to glucose and subsequently yeast strains convert glucose into ethanol. Simultaneous saccharification and fermentation is a process whereby cellulose present in biomass, e.g., corn stover, is converted to glucose and, at the same time and in the same reactor, yeast strains convert glucose into ethanol. Thus, in another approach, the variant CBH type cellulase of the invention finds utility in the degradation of biomass to ethanol. Ethanol production from readily available sources of cellulose provides a stable, renewable fuel source.

[211] Cellulose-based feedstocks are comprised of agricultural wastes, grasses and woods and other low-value biomass such as municipal waste (e.g., recycled paper, yard clippings, etc.). Ethanol may be produced from the fermentation of any of these cellulosic

feedstocks. However, the cellulose must first be converted to sugars before there can be conversion to ethanol.

[212] A large variety of feedstocks may be used with the inventive variant CBH and the one selected for use may depend on the region where the conversion is being done. For example, in the Midwestern United States agricultural wastes such as wheat straw, corn stover and bagasse may predominate while in California rice straw may predominate. However, it should be understood that any available cellulosic biomass may be used in any region.

[213] A cellulase composition containing an enhanced amount of cellobiohydrolase finds utility in ethanol production. Ethanol from this process can be further used as an octane enhancer or directly as a fuel in lieu of gasoline which is advantageous because ethanol as a fuel source is more environmentally friendly than petroleum derived products. It is known that the use of ethanol will improve air quality and possibly reduce local ozone levels and smog. Moreover, utilization of ethanol in lieu of gasoline can be of strategic importance in buffering the impact of sudden shifts in non-renewable energy and petro-chemical supplies.

[214] Ethanol can be produced via saccharification and fermentation processes from cellulosic biomass such as trees, herbaceous plants, municipal solid waste and agricultural and forestry residues. However, the ratio of individual cellulase enzymes within a naturally occurring cellulase mixture produced by a microbe may not be the most efficient for rapid conversion of cellulose in biomass to glucose. It is known that endoglucanases act to produce new cellulose chain ends which themselves are substrates for the action of cellobiohydrolases and thereby improve the efficiency of hydrolysis of the entire cellulase system. Therefore, the use of increased or optimized cellobiohydrolase activity may greatly enhance the production of ethanol.

[215] Thus, the inventive cellobiohydrolase finds use in the hydrolysis of cellulose to its sugar components. In one embodiment, a variant cellobiohydrolase is added to the biomass prior to the addition of a fermentative organism. In a second embodiment, a variant cellobiohydrolase is added to the biomass at the same time as a fermentative organism. Optionally, there may be other cellulase components present in either embodiment.

[216] In another embodiment the cellulosic feedstock may be pretreated. Pretreatment may be by elevated temperature and the addition of either of dilute acid, concentrated acid or dilute alkali solution. The pretreatment solution is added for a time sufficient to at least partially hydrolyze the hemicellulose components and then neutralized.

[217] The major product of CBHI action on cellulose is cellobiose which is available for conversion to glucose by BG activity (for instance in a fungal cellulase product). Either by the pretreatment of the cellulosic biomass or by the enzymatic action on the biomass, other sugars, in addition to glucose and cellobiose, can be made available from the biomass. The hemi-cellulose content of the biomass can be converted (by hemi-cellulases) to sugars such as xylose, galactose, mannose and arabinose. Thus, in a biomass conversion process, enzymatic saccharification can produce sugars that are made available for biological or chemical conversions to other intermediates or end-products. Therefore, the sugars generated from biomass find use in a variety of processes in addition to the generation of ethanol. Examples of such conversions are fermentation of glucose to ethanol (as reviewed by M.E. Himmel *et al.* pp2-45, in "Fuels and Chemicals from Biomass", ACS Symposium Series 666, ed B.C. Saha and J. Woodward, 1997) and other biological conversions of glucose to 2,5-diketo-D-gluconate (US Patent No. 6,599,722), lactic acid (R. Datta and S-P. Tsai pp224-236, *ibid*), succinate (R.R. Gokarn, M.A. Eiteman and J. Sridhar pp237-263, *ibid*), 1,3-propanediol (A-P. Zheng, H. Biebl and W-D. Deckwer pp264-279, *ibid*), 2,3-butanediol (C.S. Gong, N. Cao and G.T. Tsao pp280-293, *ibid*), and the chemical and biological conversions of xylose to xylitol (B.C. Saha and R.J. Bothast pp307-319, *ibid*). See also, for example, WO 98/21339.

[218] The detergent compositions of this invention may employ besides the cellulase composition (irrespective of the cellobiohydrolase content, i.e., cellobiohydrolase -free, substantially cellobiohydrolase -free, or cellobiohydrolase enhanced), a surfactant, including anionic, non-ionic and ampholytic surfactants, a hydrolase, building agents, bleaching agents, bluing agents and fluorescent dyes, caking inhibitors, solubilizers, cationic surfactants and the like. All of these components are known in the detergent art. The cellulase composition as described above can be added to the detergent composition either in a liquid diluent, in granules, in emulsions, in gels, in pastes, and the like. Such forms are well known to the skilled artisan. When a solid detergent composition is employed, the cellulase composition is preferably formulated as granules. Preferably, the granules can be formulated so as to contain a cellulase protecting agent. For a more thorough discussion, see US Patent Number 6,162,782 entitled "Detergent compositions containing cellulase compositions deficient in CBH I type components," which is incorporated herein by reference.

[219] Preferably the cellulase compositions are employed from about 0.00005 weight percent to about 5 weight percent relative to the total detergent composition. More

preferably, the cellulase compositions are employed from about 0.0002 weight percent to about 2 weight percent relative to the total detergent composition.

[220] In addition the variant *cbh I* nucleic acid sequence finds utility in the identification and characterization of related nucleic acid sequences. A number of techniques useful for determining (predicting or confirming) the function of related genes or gene products include, but are not limited to, (A) DNA/RNA analysis, such as (1) overexpression, ectopic expression, and expression in other species; (2) gene knock-out (reverse genetics, targeted knock-out, viral induced gene silencing (VIGS, see Baulcombe, 100 Years of Virology, Calisher and Horzinek eds., Springer-Verlag, New York, NY 15:189-201, 1999); (3) analysis of the methylation status of the gene, especially flanking regulatory regions; and (4) in situ hybridization; (B) gene product analysis such as (1) recombinant protein expression; (2) antisera production, (3) immunolocalization; (4) biochemical assays for catalytic or other activity; (5) phosphorylation status; and (6) interaction with other proteins via yeast two-hybrid analysis; (C) pathway analysis, such as placing a gene or gene product within a particular biochemical or signaling pathway based on its overexpression phenotype or by sequence homology with related genes; and (D) other analyses which may also be performed to determine or confirm the participation of the isolated gene and its product in a particular metabolic or signaling pathway, and help determine gene function.

[221] All patents, patent applications, articles and publications mentioned herein, are hereby expressly incorporated herein by reference.

EXAMPLES

[222] The present invention is described in further detain in the following examples which are not in any way intended to limit the scope of the invention as claimed. The attached Figures are meant to be considered as integral parts of the specification and description of the invention. All references cited are herein specifically incorporated by reference for all that is described therein.

EXAMPLE 1
Alignment of known Cel7A cellulases

[223] The choice of several of the mutations was determined by first aligning *Hypocrea jecorina* Cel7A to its 41 family members using structural information and a modeling program. The alignment of the primary amino acid sequence of all 42 family members is shown in Figure 8.

[224] For four of the members (i.e., 20VW.1, 1A39, 6CEL and 1EG1.1), the crystal structure had been previously determined. The 4 aligned proteins for which there were published structures had their alignment locked for all residues whose backbone atoms were within a specific RMS deviation (RMS less than or equal to 2.0 A). The tertiary structural alignment of the four sequences was performed using MOE version 2001.01 by Chemical Computing Group, Montreal Canada. The overlapping structural elements were used to freeze the primary structures of the four sequences. The remaining 38 sequences then had their primary amino acid structure aligned with the frozen four using MOE with secondary structure prediction on and other parameters set to their default settings.

[225] Based on the alignments, various single and multiple amino acid mutations were made in the protein by site mutagenesis.

[226] Single amino acid mutations were based on the following rationale (see also Table 1): After examining the conservation of amino acids between the homologues, sites were picked in the *H. jecorina* sequence where a statistical preference for another amino acid was seen amongst the other 41 sequences (e.g.: at position 77 the Ala, only present in *H. jecorina* and 3 other homologues, was changed to Asp, present in 22 others). The effect of each substitution on the structure was then modeled.

Table 1: Cel7A Variants and Rationale for Change

Cel7A Variants and Rationale for Change		Tm	ΔTm
Wild Type *H. jecorina*		62.5	
(4)A77D(22)	3 possible H-bonds to Q7 and I80	62.2	-0.3
(7)S113D(18)	numerous new H-bonds to backbone to stabilize turn	62.8	0.3
(8)L225F(13)	better internal packing	61.6	-0.9
(5)L288F(17)	better internal packing	62.4	-0.1
(1)A299E(24)	extra ligand to cobalt atom observed in crystal structure	61.2	-1.3
(4)N301K(11)	salt bridges to E295 and E325	63.5	1.0
(5)T356L(20)	better internal packing	62.6	0.1
(2)G430F(17)	better surface packing	61.7	-0.8

[227] Multiple amino acid mutations were based on a desire to affect the stability, processivity, and product inhibition of the enzyme. The following multiple site changes in the *H. jecorina* sequence were constructed:

1) Thr 246 Cys + Tyr 371 Cys
2) Thr 246 Ala + Arg 251 Ala + Tyr 252 Ala
3) Thr 380 Gly + Tyr 381 Asp + Arg 394 Ala + deletion of Residues 382 to 393, inclusive
4) Thr 380 Gly + Tyr 381 Asp + Arg 394 Ala
5) Tyr 252 Gln + Asp 259 Trp + Ser 342 Tyr

[228] The T246A/R251A/Y252A and the other triple + deletion mutant are both predicted to decrease the product inhibition of the enzyme. The Thr246Cys + Tyr371Cys is predicted to increase the stability of the enzyme and increase the processitivity of it. The D259W/Y252Q/S342Y variant is predicted to affect the product inhibition of the enzyme.

[229] Other single and multiple mutations were constructed using methods well known in the art (see references above) and are presented in Table 2.

Table 2: *H. jecorina* CBH I variants

Mutations
S8P
N49S
A68T
A77D
N89D
S92T
S113N
S113D
L225F
P227A
P227L
D249K
T255P
D257E
S279N
L288F
E295K
S297T
A299E
N301K
T332K
T332Y
T332H
T356L
F338Y
V393G
G430F
T41I (plus deletion of Thr @ 445)
V403D/T462I
S196T/S411F
E295K/S398T
A112E/T226A
T246C/Y371C
G22D/S278P/T296P
S8P/N103I/S113N
S113T/T255P/K286M
P227L/E325K/Q487L
P227T/T484S/F352L
T246A/R251A/Y252A

Mutations
T380G/Y381D/R394A
Y252Q/D259W/S342Y
A68T/G440R/P491L
Q17L/E193V/M213I/F352L
S8P/N49S/A68T/S113N
A112E/P227L/S278P/T296P
S8P/N49S/A68T/N103I/S113N
S8P/N49S/A68T/S278P/T296P
G22D/N49S/A68T/S278P/T296P
G22D/N103I/S113N/S278P/T296P
S8P/N49S/A68T/S113N/P227L
S8P/N49S/A68T/A112E/T226A
S8P/N49S/A68T/A112E/P227L
T41I/A112E/P227L/S278P/T296P
S8P/T41I/N49S/A68T/S113N/P227L
S8P/T41I/N49S/A68T/A112E/P227L
G22D/N49S/A68T/P227L/S278P/T296P
G22D/N49S/A68T/N103I/S113N/S278P/T296P
G22D/N49S/A68T/N103I/S113N/P227L/S278P/ T296P
G22D/N49S/A68T/N103I/A112E/P227L/S278P/ T296P
G22D/N49S/N64D/A68T/N103I/S113N/S278P/ T296P
S8P/T41I/N49S/A68T/S92T/S113N/P227L/D249K/S411F
S8P/G22D/T41I/N49S/A68T/N103I/S113N/S278P/T296P
S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/S278P/T296P
S8P/G22D/T41I/N49S/A68T/S113N/P227L/D249K/S278P/N301R
S8P/G22D/T41I/N49S/A68T/S92T/S113N/P227L/D249K/S411F
S8P/T41I/N49S/A68T/S92T/S113N/P227L/D249K/V403D/T462I
S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/D249K/S278P/T296P
S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/S278P/T296P/N301R
S8P/G22D/T41I/N49S/A68T/S113N/P227L/D249K/S278P/T296P/N301R
S8P/T41I/N49S/S57N/A68T/S113N/P227L/D249K/S278P/T296P/N301R
S8P/G22D/T41I/N49S/A68T/S92T/S113N/P227L/D249K/V403D/T462I
S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/D249K/S278P/T296P/N301R
S8P/G22D/T41I/N49S/A68T/S92T/S113N/S196T/P227L/D249K/T255P/S278P/T296P/N301R/ E325K/S411F
S8P/T41I/N49S/A68T/S92T/S113N/S196T/P227L/D249K/T255P/S278P/T296P/N301R/E325K/ V403D/S411F/T462I
S8P/G22D/T41I/N49S/A68T/S92T/S113N/S196T/P227L/D249K/T255P/S278P/T296P/N301R/ E325K/V403D/S411F/T462I

EXAMPLE 2

Cloning and Expression of CBHI variants in *H. jecorina*

A. Construction of the *H. jecorina* general-purpose expression plasmid-PTEX.

[230] The plasmid, pTEX was constructed following the methods of Sambrook et al. (1989), *supra*, and is illustrated in FIG. 7. This plasmid has been designed as a multi-purpose expression vector for use in the filamentous fungus *Trichoderma longibrachiatum*.

The expression cassette has several unique features that make it useful for this function. Transcription is regulated using the strong CBH I gene promoter and terminator sequences for *T. longibrachiatum*. Between the CBHI promoter and terminator there are unique PmeI and SstI restriction sites that are used to insert the gene to be expressed. The *T. longibrachiatum* pyr4 selectable marker gene has been inserted into the CBHI terminator and the whole expression cassette (CBHI promoter-insertion sites-CBHI terminator-pyr4 gene-CBHI terminator) can be excised utilizing the unique NotI restriction site or the unique NotI and NheI restriction sites.

[231] This vector is based on the bacterial vector, pSL1180 (Pharmacia Inc., Piscataway, N.J.), which is a PUC-type vector with an extended multiple cloning site. One skilled in the art would be able to construct this vector based on the flow diagram illustrated in FIG. 7.

[232] The vector pTrex2L was constructed from pTrex2, a derivative of pTEX. The sequence for pTrex2 is given in Figure 6.

[233] The exact plasmid used is not that important as long as the variant protein is expressed at a useful level. However, maximizing the expression level by forcing integration at the cbh1 locus is advantageous.

B. Cloning

[234] Using methods known in the art a skilled person can clone the desired CBH I variant into an appropriate vector. As noted above, the exact plasmid used is not that important as long as the variant protein is expressed at a useful level. The following description of the preparation of one of the inventive variant CBH I enzymes can be utilized to prepare any of the inventive variants described herein.

[235] The variant *cbh 1* genes were cloned into the pTrex2L vector.

[236] Construction of plasmid pTrex2L was done as follows: The 6 nucleotides between the unique Sac II and Asc I sites of pTrex2 were replaced with a synthetic linker containing a BstE II and BamH I sites to produce plasmid Trex2L. The complementary synthetic linkers

21-mer synthetic oligo CBHlink1+: GGTTT**GGATCCGGTCACC**AGG

and

27-mer synthetic oligo CBHlink-: CGCGCCT**GGTGACCGGATCC**AAACCGC

were annealed.

[237] The pTrex2 was digested with Sac II and Asc I. The annealed linker was then ligated into pTrex2 to create pTrex2L. The plasmid was then digested with an appropriate restriction enzyme(s) and a wild type CBH I gene was ligated into the plasmid.

[238] Primers were used to introduce the desired mutations into the wild-type gene. It will be understood that any method that results in the introduction of a desired alteration or mutation in the gene may be used. Synthetic DNA primers were used as PCR templates for mutant constructions. It is well within the knowledge of the skilled artisan to design the primers based on the desired mutation to be introduced.

[239] The mutagenic templates were extended and made double stranded by PCR using the synthetic DNA oligonucleotides. After 25 PCR cycles the final product was primarily a 58 bp double stranded product comprising the desired mutation. The mutagenic fragments were subsequently attached to wild-type CBH I fragments and ligated into the plasmid using standard techniques.

C. Transformation and Expression

[240] The prepared vector for the desired variant was transformed into the uridine auxotroph version of the double or quad deleted Trichoderma strains (see Table 3; see also U.S. Patent Nos. 5,861,271 and 5,650,322) and stable transformants were identified.

Table 3: Transformation/Expression strain

CBH I Variant	**Expression Strain**
A77D	quad-delete strain (1A52)
S113D	double-delete strain
L225F	double-delete strain
L288F	double-delete strain
A299E	quad-delete strain (1A52)
N301K	quad-delete strain (1A52)
T356L	double-delete strain
G430F	quad-delete strain (1A52)
T246C/Y371C	quad-delete strain (1A52)
T246A/R251A/Y252A	quad-delete strain (1A52)
Y252Q/D259W/S342Y	quad-delete strain (1A52)
T380G/Y381D/R394A	quad-delete strain (1A52)
T380G/Y381D/R394A plus deletion of 382-393	quad-delete strain (1A52)

"double-delete" (Δ CBHI & Δ CBHII) and the "quad-delete" (Δ CBHI & Δ CBHII, Δ EGI & Δ EGII) T.reesei host strains

[241] To select which transformants expressed variant CBH I, DNA was isolated from strains following growth on Vogels+1% glucose and Southern blot experiments performed using an isolated DNA fragment containing only the variant CBH I. Transformants were isolated having a copy of the variant CBH I expression cassette integrated into the genome of the host cell. Total mRNA was isolated from the strains following growth for 1 day on Vogels+1% lactose. The mRNA was subjected to Northern analysis using the variant CBH I coding region as a probe. Transformants expressing variant CBH I mRNA were identified.

[242] One may obtain any other novel variant CBH I cellulases or derivative thereof by employing the methods described above.

EXAMPLE 3

Expression of CBH1 variants in *A. niger*

[243] The PCR fragments were obtained using the following primers and protocols

[244] The following DNA primers were constructed for use in amplification of homologous CBH1 genes from genomic DNA's isolated from various microorganisms. All symbols used herein for protein and DNA sequences correspond to IUPAC IUB Biochemical Nomenclature Commission codes.

[245] Homologous 5' (FRG192) and 3' (FRG193) primers were developed based on the sequence of CBH1 from *Trichoderma reesei*. Both primers contained Gateway cloning sequences from Invitrogen® at the 5' of the primer. Primer FRG192 contained attB1 sequence and primer FRG193 contained attB2 sequence.

Sequence of FRG192 without the attB1:
ATGTATCGGAAGTTGGCCG (signal sequence of CBH1 *H. jecorina*) (SEQ ID NO: 3)

Sequence of FRG193 without the attB2:
TTACAGGCACTGAGAGTAG (cellulose binding module of CBH1 *H. jecorina*) (SEQ ID NO: 4)

[246] The *H. jecorina* CBH I cDNA clone served as template.

[247] PCR conditions were as follows: 10 µL of 10X reaction buffer (10X reaction buffer comprising 100mM Tris HCl, pH 8-8.5; 250 mM KCl; 50 mM $(NH_4)_2SO_4$; 20 mM $MgSO_4$); 0.2 mM each of dATP, dTTP, dGTP, dCTP (final concentration), 1 µL of 100 ng/µL genomic DNA, 0.5 µL of PWO polymerase (Boehringer Mannheim, Cat # 1644-947) at 1 unit per µL, 0.2µM of each primer, FRG192 and FRG193, (final concentration), 4µl DMSO and water to 100 µL.

[248] Various sites in *H. jecorina* CBH1 may be involved in the thermostability of the variants and the *H. jecorina* CBH1 gene was therefore subjected to mutagenesis.

[249] The fragments encoding the variants were purified from an agarose gel using the Qiagen Gel extraction KIT. The purified fragments were used to perform a clonase reaction with the pDONR™201 vector from Invitrogen® using the Gateway™ Technology instruction manual (version C) from Invitrogen®, hereby incorporated by reference herein. Genes were then transferred from this ENTRY vector to the destination vector (pRAXdes2) to obtain the expression vector pRAXCBH1.

[250] Cells were transformed with an expression vector comprising a variant CBH I cellulase encoding nucleic acid. The constructs were transformed into *A. niger var. awamori* according to the method described by Cao *et al* (Cao Q-N, Stubbs M, Ngo KQP, Ward M, Cunningham A, Pai EF, Tu G-C and Hofmann T (2000) Penicillopepsin-JT2 a recombinant enzyme from *Penicillium janthinellum* and contribution of a hydrogen bond in subsite *S*3 to *kcat Protein Science* 9:991-1001).

[251] Transformants were streaked on minimal medium plates (Ballance DJ, Buxton FP, and Turner G (1983) Transformation of *Aspergillus nidulans* by the orotidine-5'-phosphate decarboxylase gene of *Neurospora crassa Biochem Biophys Res Commun* 112:284-289) and grown for 4 days at 30°C. Spores were collected using methods well known in the art (See <http://www.fgsc.net/fgn48/Kaminskyj.htm>). *A. nidulans* conidia are harvested in water (by rubbing the surface of a conidiating culture with a sterile bent glass rod to dislodge the spores) and can be stored for weeks to months at 4°C without a serious loss of viability. However, freshly harvested spores germinate more reproducibly. For long-term storage, spores can be stored in 50% glycerol at –20°C, or in 15-20% glycerol at –80°C. Glycerol is more easily pipetted as an 80% solution in water. 800µl of aqueous conidial suspension (as made for 4°C storage) added to 200µl 80% glycerol is used for a –80°C stock; 400 µl suspension added to 600 µl 80% glycerol is used for a –20°C stock. Vortex before freezing. For mutant collections, small pieces of conidiating cultures can be excised and placed in 20% glycerol, vortexed, and frozen as –80°C stocks. In our case we store them in 50% glycerol at –80°C.

[252] *A. niger* var *awamori* transformants were grown on minimal medium lacking uridine (Ballance et al. 1983). Transformants were screened for cellulase activity by inoculating 1cm^2 of spore suspension from the sporulated grown agar plate into 100ml shake flasks for 3 days at 37°C as described by Cao et al. (2000).

[253] The CBHI activity assay is based on the hydrolysis of the nonfluorescent 4-methylumbelliferyl-ß-lactoside to the products lactose and 7-hydroxy-4-methylcoumarin,

the latter product is responsible for the fluorescent signal. Pipette 170 µl 50 mM NaAc buffer pH 4.5 in a 96-well microtiter plate (MTP) (Greiner, Fluotrac 200, art. nr. 655076) suitable for fluorescence. Add 10 µl of supernatant and then add 10 µl of MUL (1 mM 4-methylumbelliferyl-ß-lactoside (MUL) in milliQ water) and put the MTP in the Fluostar Galaxy (BMG Labtechnologies; D-77656 Offenburg). Measure the kinetics for 16 min. (8 cycles of 120s each) using $\lambda_{320\ nm}$ (excitation) and $\lambda_{460\ nm}$ (emission) at 50°C. Supernatents having CBH activity were then subjected to Hydrophobic Interaction Chromatography.

EXAMPLE 4

Stability of CBH 1 variants

[254] CBH I cellulase variants were cloned and expressed as above (see Examples 2 and 3). Cel7A wild type and variants were then purified from cell-free supernatants of these cultures by column chromatography. Proteins were purified using hydrophobic interaction chromatography (HIC). Columns were run on a BioCAD® Sprint Perfusion Chromatography System using Poros® 20 HP2 resin both made by Applied Biosystems.

[255] HIC columns were equilibrated with 5 column volumes of 0.020 M sodium phosphate, 0.5 M ammonium sulfate at pH 6.8. Ammonium sulfate was added to the supernatants to a final concentration of approximately 0.5 M and the pH was adjusted to 6.8. After filtration, the supernatant was loaded onto the column. After loading, the column was washed with 10 column volumes of equilibration buffer and then eluted with a 10 column volume gradient from 0.5 M ammonium sulfate to zero ammonium sulfate in 0.02 M sodium phosphate pH 6.8. Cel7A eluted approximately mid-gradient. Fractions were collected and pooled on the basis of reduced, SDS-PAGE gel analysis.

[256] The melting points were determined according to the methods of Luo, *et al., Biochemistry* **34**:10669 and Gloss, *et al., Biochemistry* **36**:5612. See also Sandgren at al. (2003) Protein Science 12(4) pp848.

[257] Data was collected on the Aviv 215 circular dichroism spectrophotometer. The native spectra of the variants between 210 and 260 nanometers were taken at 25°C. Buffer conditions were 50 mM Bis Tris Propane/50 mM ammonium acetate/glacial acetic acid at pH 5.5. The protein concentration was kept between 0.25 and 0.5 mgs/mL. After determining the optimal wavelength to monitor unfolding, the samples were thermally denatured by ramping the temperature from 25°C to 75°C under the same buffer conditions. Data was collected for 5 seconds every 2 degrees. Partially reversible unfolding was monitored at 230 nanometers in a 0.1 centimeter path length cell. While at 75°C, an unfolded spectra was collected as described above. The sample was then

cooled to 25°C to collect a refolded spectra. The difference between the three spectra at 230nm was used to assess the variants reversibility.

[258] The thermal denaturation profiles are shown in Figure 9A and 9B for wildtype CBH I and various variant CBH I's. See also Table 4.

Table 4: Thermal Stability of Variant CBH I cellulases

H. jecorina CBH I Residue Substitution	Tm	delta Tm	% rev 230nm
Wild type	**62.5**		23
S8P	63.1	0.6	
N49S	63.7	1.2	
A68T	63.7	1.2	32
A77D	62.2	-0.3	
N89D	63.6	1.1	50
S92T	64.4	1.9	25
S113D	62.8	0.3	
S113N	64.0	1.5	
L225F	61.6	-0.9	
P227A	64.8	2.3	49
P227L	65.2	2.7	45
D249K	64.0	1.5	39
T255P	64.4	1.9	35
S279N	62.4	-0.1	~95
E295K	64.0	1.5	~95
T332K	63.3	0.8	37
T332Y	63.3	0.8	37
T332H	62.7	0.2	64
F338Y	60.8	-1.7	~95
G430F	61.7	-0.8	
L288F	62.4	-0.1	
A299E	61.2	-1.3	
N301K	63.5	1.0	
T356L	62.6	0.1	
D257E	61.8	-0.7	45
V393G	61.7	-0.8	43
S297T	63.3	0.8	31
T41I plus deletion @ T445	64.2	1.7	
T246C/Y371C	65.0	2.5	
S196T/S411F	65.3	2.8	27
E295K/S398T	63.9	1.4	36
V403D/T462I	64.5	2	53
A112E/T226A	63.5	1.0	
A68T/G440R/P491L	63.1	0.6	32
G22D/S278P/T296P	63.6	1.1	
T246A/R251A/Y252A	63.5	1.0	
T380G/Y381D/R394A	58.1	-4.4	
Y252Q/D259W/S342Y	59.9	-2.6	50
S113T/T255P/K286M	63.8	1.3	16

H. jecorina CBH I Residue Substitution	Tm	delta Tm	% rev 230nm
P227L/E325K/Q487L	64.5	2.0	22
P227T/T484S/F352L	64.2	1.7	45
Q17L/E193V/M213I/F352L	64.0	1.5	34
S8P/N49S/A68T/S113N	64.5	2.0	90
S8P/N49S/A68T/S113N/P227L	66.0	3.5	86
T41I/A112E/P227L/S278P/T296P	66.1	3.6	48
S8P/N49S/A68T/A112E/T226A	64.6	2.1	46
S8P/N49S/A68T/A112E/P227L	65.2	2.7	32
S8P/T41I/N49S/A68T/A112E/P227L	67.6	5.1	40
G22D/N49S/A68T/P227L/S278P/T296P	65.9	3.4	26
G22D/N49S/A68T/N103I/S113N/P227L/S278P/T296P	65.3	2.8	72
G22D/N49S/A68T/N103I/A112E/P227L/S278P/T296P	65.1	2.6	20
G22D/N49S/N64D/A68T/N103I/S113N/S278P/T296P	61.4	-1.1	75
S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/S278P/ T296P	68.8	6.3	56
S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/D249K/ S278P/T296P	69.0	6.5	71
S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/S278P/ T296P/N301R	68.7	6.2	70
S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/D249K/ S278P/T296P/N301R	68.8	6.3	74
S8P/G22D/T41I/N49S/A68T/S113N/P227L/D249K/S278P/ T296P/N301R	69.9	7.4	88
S8P/T41I/N49S/S57N/A68T/S113N/P227L/D249K/S278P/ T296P/N301R	68.9	6.4	~100
S8P/G22D/T41I/N49S/A68T/S113N/P227L/D249K/S278P/ N301R	68.7	6.2	92
S8P/T41I/N49S/A68T/S92T/S113N/P227L/D249K/V403D/ T462I	68.8	6.3	~100
S8P/G22D/T41I/N49S/A68T/S92T/S113N/P227L/D249K/ V403D/T462I	68.5	6.0	~100
S8P/T41I/N49S/A68T/S92T/S113N/P227L/D249K/S411F	68.6	6.1	~100
S8P/G22D/T41I/N49S/A68T/S92T/S113N/P227L/D249K/ S411F	69.5	7.0	~100
S8P/G22D/T41I/N49S/A68T/S92T/S113N/S196T/P227L/ D249K/T255P/S278P/T296P/N301R/E325K/S411F	70.7	8.2	~100
S8P/T41I/N49S/A68T/S92T/S113N/S196T/P227L/D249K/ T255P/S278P/T296P/N301R/E325K/V403D/S411F/T462I	71.0	8.5	~100
S8P/G22D/T41I/N49S/A68T/S92T/S113N/S196T/P227L/ D249K/T255P/S278P/T296P/N301R/E325K/V403D/S411F/ T462I	70.9	8.4	~100

[259] Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed

should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the claims.

CLAIMS

1. A variant CBH I cellulase, wherein said variant comprises a substitution or deletion at a position corresponding to one or more of residues S8, Q17, G22, T41, N49, S57, N64, A68, A77, N89, S92, N103, A112, S113, E193, S196, M213, L225, T226, P227, T246, D249, R251, Y252, T255, D257, D259, S278, S279, K286, L288, E295, T296, S297, A299, N301, E325, T332, F338, S342, F352, T356, Y371, T380, Y381, V393, R394, S398, V403, S411, G430, G440, T445, T462, T484, Q487, and P491 in CBH I from *Hypocrea jecorina* (SEQ ID NO: 2).

2. A variant CBH I cellulase according to Claim 1, wherein said variant comprises a substitution at a position corresponding to one or more of residues S8P, Q17L, G22D, T41I, N49S, S57N, N64D, A68T, A77D, N89D, S92T, N103I, A112E, S113(T/N/D), E193V, S196T, M213I, L225F, T226A, P227(L/T/A), T246(C/A), D249K, R251A, Y252(A/Q), T255P, D257E, D259W, S278P, S279N, K286M, L288F, E295K, T296P, S297T, A299E, N301(R/K), E325K, T332(K/Y/H), F338Y, S342Y, F352L, T356L, Y371C, T380G, Y381D, V393G, R394A, S398T, V403D, S411F, G430F, G440R, T462I, T484S, Q487L and/or P491L in CBH I from *Hypocrea jecorina* (SEQ ID NO: 2).

3. A variant CBH I cellulose according to Claim 2, further comprising a deletion at a position corresponding to T445 in CBH I from *Hypocrea jecorina* (SEQ ID NO: 2).

4. A variant CBH I cellulase, wherein said variant comprises a substitution at a position corresponding to a residue selected from the group consisting of S8P, N49S, A68T, A77D, N89D, S92T, S113(N/D), L225F, P227(A/L/T), D249K, T255P, D257E, S279N, L288F, E295K, S297T, A299E, N301(R/K), T332(K/Y/H), F338Y, T356L, V393G, G430F in CBH I from *Hypocrea jecorina* (SEQ ID NO: 2).

5. A variant CBH I cellulase, wherein said variant CBH I consists essentially of the mutations selected from the group consisting of

xl. A112E/T226A;

xli. S196T/S411F;

xlii. E295K/S398T;

xliii. T246C/Y371C;

xliv. V403D/T462I

xlv. T41I plus deletion at T445

xlvi. A68T/G440R/P491L;

xlvii. G22D/S278P/T296P;

xlviii. T246A/R251A/Y252A;

xlix. T380G/Y381D/R394A;

l. Y252Q/D259W/S342Y;

li. S113T/T255P/K286M;

lii. P227L/E325K/Q487L;

liii. P227T/T484S/F352L;

liv. Q17L/E193V/M213I/F352L;

lv. S8P/N49S/A68T/S113N;

lvi. S8P/N49S/A68T/S113N/P227L;

lvii. T41I/A112E/P227L/S278P/T296P;

lviii. S8P/N49S/A68T/A112E/T226A;

lix. S8P/N49S/A68T/A112E/P227L;

lx. S8P/T41I/N49S/A68T/A112E/P227L;

lxi. G22D/N49S/A68T/P227L/S278P/T296P;

lxii. S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/S278P/T296P;

lxiii. G22D/N49S/A68T/N103I/S113N/P227L/S278P/T296P;

lxiv. G22D/N49S/A68T/N103I/A112E/P227L/S278P/T296P;

lxv. G22D/N49S/N64D/A68T/N103I/S113N/S278P/T296P;

lxvi. S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/D249K/S278P/T296P;

lxvii. S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/S278P/T296P/N301R;

lxviii. S8P/G22D/T41I/N49S/A68T/N103I/S113N/P227L/D249K/S278P/T296P/N301R

lxix. S8P/G22D/T41I/N49S/A68T/S113N/P227L/D249K/S278P/T296P/N301R;

lxx. S8P/T41I/N49S/S57N/A68T/S113N/P227L/D249K/S278P/T296P/N301R;

lxxi. S8P/G22D/T41I/N49S/A68T/S113N/P227L/D249K/S278P/N301R;

lxxii. S8P/T41I/N49S/A68T/S92T/S113N/P227L/D249K/V403D/T462I;

lxxiii. S8P/G22D/T41I/N49S/A68T/S92T/S113N/P227L/D249K/V403D/T462I;

lxxiv. S8P/T41I/N49S/A68T/S92T/S113N/P227L/D249K/S411F;

lxxv. S8P/G22D/T41I/N49S/A68T/S92T/S113N/P227L/D249K/S411F;

lxxvi. S8P/G22D/T41I/N49S/A68T/S92T/S113N/S196T/P227L/D249K/T255P/S278P/T296P/N301R/E325K/S411F;

lxxvii. S8P/T41I/N49S/A68T/S92T/S113N/S196T/P227L/D249K/T255P/S278P /T296P/N301R/E325K/V403D/S411F/T462I;

lxxviii. S8P/G22D/T41I/N49S/A68T/S92T/S113N/S196T/P227L/D249K/T255P/ S278P/T296P/N301R/E325K/V403D/S411F/T462I

in CBH I from *Hypocrea jecorina* (SEQ ID NO:2).

6. A nucleic acid encoding a CBH I variant according to claim 1.

7. A nucleic acid encoding a CBH I variant according to claim 4.

8. A nucleic acid encoding a CBH I variant according to claim 5.

9. A vector comprising a nucleic acid encoding a CBH I variant of claim 6.

10. A vector comprising a nucleic acid encoding a CBH I variant of claim 7.

11. A vector comprising a nucleic acid encoding a CBH I variant of claim 8.

12. A host cell transformed with the vector of claim 9.

13. A host cell transformed with the vector of claim 10.

14. A host cell transformed with the vector of claim 11.

15. A method of producing a CBH I variant comprising the steps of:

(a) culturing the host cell according to claim 12 in a suitable culture medium under suitable conditions to produce CBH I variant;
(b) obtaining said produced CBH I variant.

16. A method of producing a CBH I variant comprising the steps of:

(a) culturing the host cell according to claim 13 in a suitable culture medium under suitable conditions to produce CBH I variant;
(b) obtaining said produced CBH I variant.

17. A method of producing a CBH I variant comprising the steps of:

(a) culturing the host cell according to claim 14 in a suitable culture medium under suitable conditions to produce CBH I variant;
(b) obtaining said produced CBH I variant.

18. A detergent composition comprising a surfactant and a CBH I variant, wherein said CBH I variant comprises a CBH I variant according to claim 1.

19. The detergent according to claim 18, wherein said detergent is a laundry detergent.

20. The detergent according to claim 18, wherein said detergent is a dish detergent.

21. A feed additive comprising a CBH I variant according to claim 1.

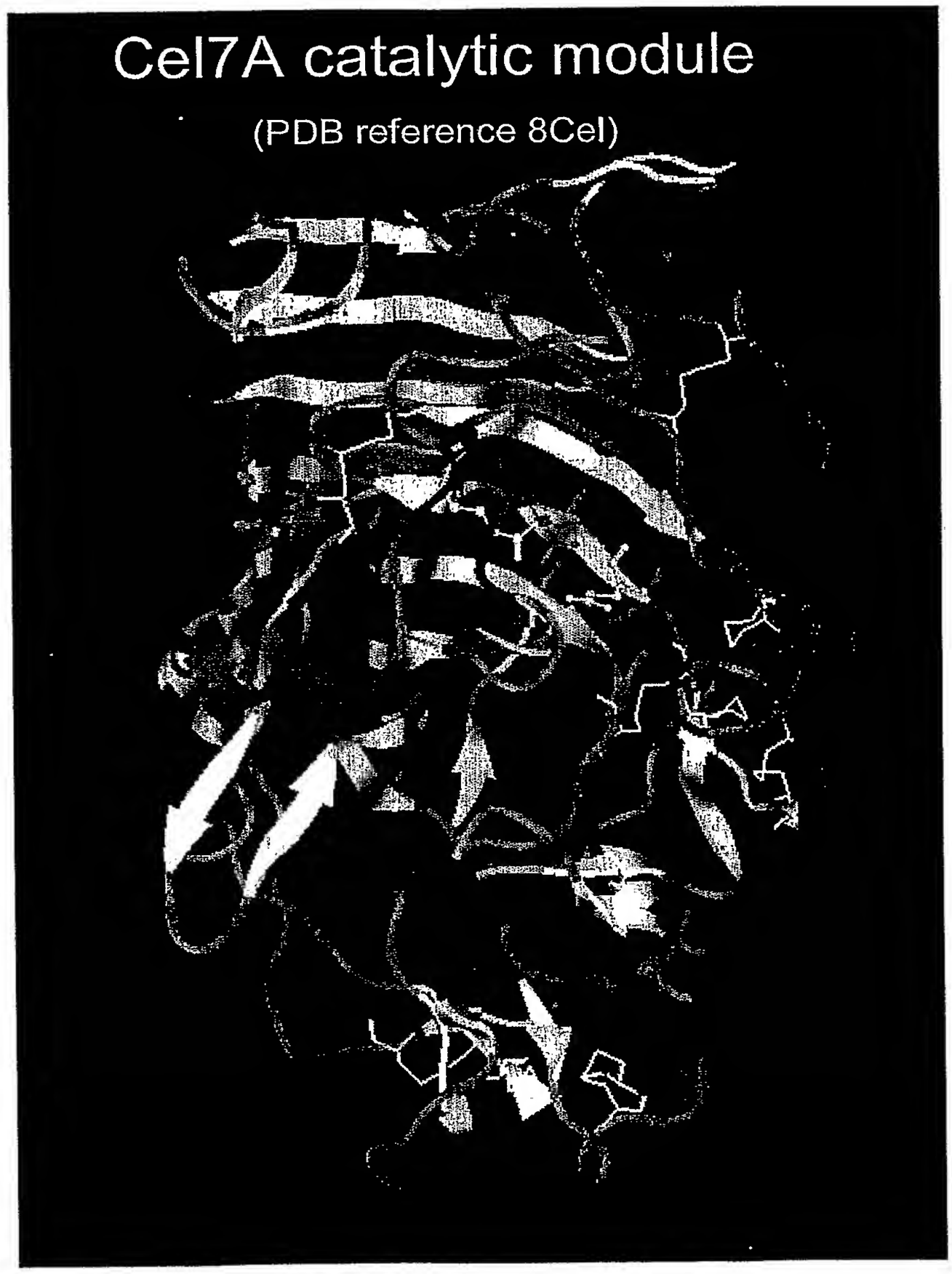
22. A method of treating wood pulp comprising contacting said wood pulp with a CBH I variant according to claim 1.

23. A method of converting biomass to sugars comprising contacting said biomass with a CBH I variant according to claim 1.

Figure 1: **Amino Acid and Nucleic Acid Sequences of *Hypocrea jecorina* Cel7A**

```
      GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle·
1     CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCCGCCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTTGCACTCA ACAGACAGGC TCCGTGGTCA

      ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
101   TCGACGCCAA CTGGCGCTGG ACTCACGCTA CGAACAGCAG CACGAACTGC TACGATGGCA ACACTTGGAG CTCGACCCTA TGTCCTGACA ACGAGACCTG

      ·AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
201   CGCGAAGAAC TGCTGTCTGG ACGGTGCCGC CTACGCGTCC ACGTACGGAG TTACCACGAG CGGTAACAGC CTCTCCATTG GCTTTGTCAC CCAGTCTGCG

      GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
301   CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGAGCG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTTCGAT GTTGATGTTT

      ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
401   CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCCACC AACACCGCTG GCGCCAAGTA

      ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
501   CGGCACGGGG TACTGTGACA GCCAGTGTCC CCGCGATCTG AAGTTCATCA ATGGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGCGAAC

      ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly·
601   ACGGGCATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTCTTACCCC CCACCCTTGC ACGACTGTCG

      ·GGlnGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgTyrGlyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyrArg·
701   GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGAACTTA CTCCGATAAC AGATATGGCG GCACTTGCGA TCCCGATGGC TGCGACTGGA ACCCATACCG

      ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaIle
801   CCTGGGCAAC ACCAGCTTCT ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGCCATC

      AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
901   AACCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

      ·GGluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp·
1001  AGGAGGCAGA ATTCGGCGGA TCCTCTTTCT CAGACAAGGG CGGCCTGACT CAGTTCAAGA AGGCTACCTC TGGCGGCATG GTTCTGGTCA TGAGTCTGTG

      ·AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerThr TyrProThrAsn GluThrSer SerThrPro GlyAlaValArg GlySerCys SerThrSer
1101  GGATGATTAC TACGCCAACA TGCTGTGGCT GGACTCCACC TACCCGACAA ACGAGACCTC CTCCACACCC GGTGCCGTGC GCGGAAGCTG CTCCACCAGC

      SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrGly AsnProSerGly·
1201  TCCGGTGTCC CTGCTCAGGT CGAATCTCAG TCTCCCAACG CCAAGGTCAC CTTCTCCAAC ATCAAGTTCG GACCCATTGG CAGCACCGGC AACCCTAGCG

      ·GGlyAsnPro ProGlyGly AsnProProGly ThrThrThr ThrArgArg ProAlaThrThr ThrGlySer SerProGly ProThrGlnSer HisTyrGly·
1301  GCGGCAACCC TCCCGGCGGA AACCCGCCTG GCACCACCAC CACCCGCCGC CCAGCCACTA CCACTGGAAG CTCTCCCGGA CCTACCCAGT CTCACTACGG

      ·GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaSerGlyThr ThrCysGln ValLeuAsn ProTyrTyrSer GlnCysLeu
1401  CCAGTGCGGC GGTATTGGCT ACAGCGGCCC CACGGTCTGC GCCAGCGGCA CAACTTGCCA GGTCCTGAAC CCTTACTACT CTCAGTGCCT G
```

Figure 2



Hypocrea jecorina **Cel7A**

497 amino acids

1 -431 in the catalytic module

432-461 in the linker region

462-497 in the cellulose binding module

12 disulfide bonds--10 in the catalytic module

E212 and E217 are the active site residues

Figure 3

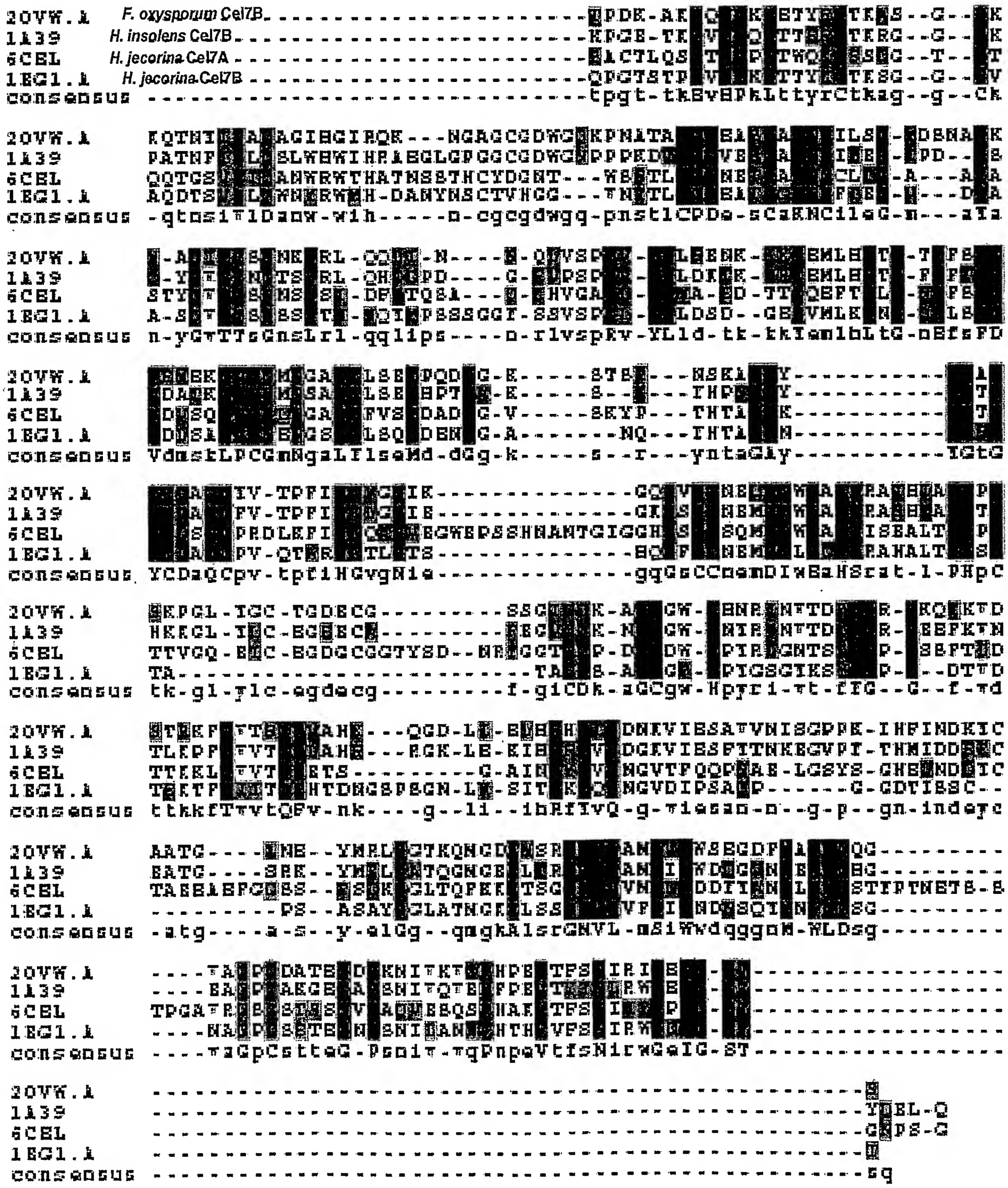
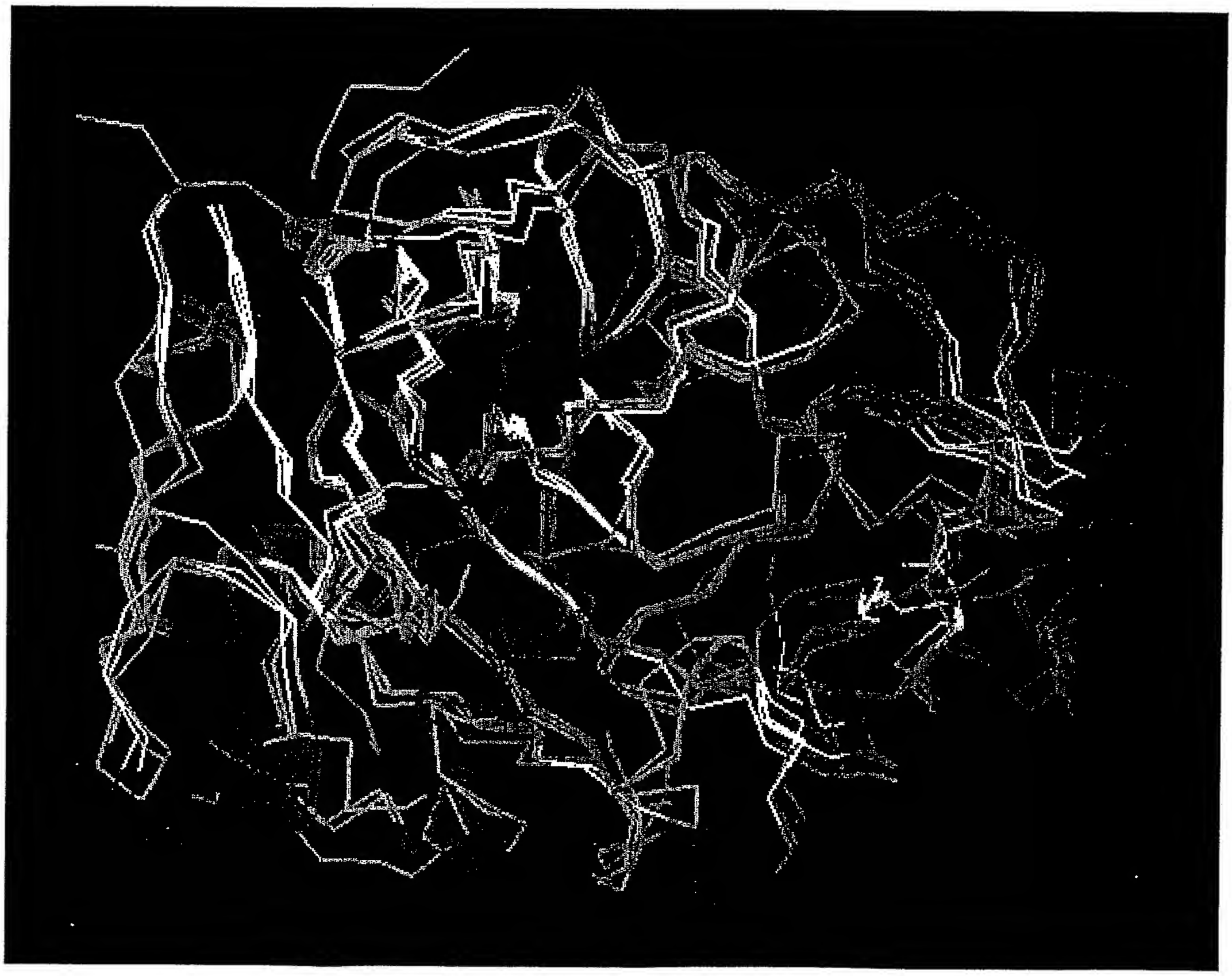

2OVW.A F. oxysporum Cel7B
1A39 H. insolens Cel7B
6CEL H. jecorina Cel7A
1EG1.A H. jecorina Cel7B
consensus

Figure 4



C-α trace of the crystal structures from the catalytic domains of four Cel7 homologues aligned and overlaid as described.

Red = α-helix, Cyan = disordered, Blue/Green = turns

Figure 5A: AminoAcid and Nucleic Acid Sequence of *Hypocrea jecorina* Cel7A mutant L225F

```
      GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle·
    1 CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCCGCCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTTGCACTCA ACAGACAGGC TCCGTGGTCA

      ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
  101 TCGACGCCAA CTGGCGCTGG ACTCACGCTA CGAACAGCAG CACGAACTGC TACGATGGCA ACACTTGGAG CTCGACCCTA TGTCCTGACA ACGAGACCTG

      ·AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
  201 CGCGAAGAAC TGCTGTCTGG ACGGTGCCGC CTACGCGTCC ACGTACGGAG TTACCACGAG CGGTAACAGC CTCTCCATTG GCTTTGTCAC CCAGTCTGCG

      GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
  301 CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGAGCG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTTCGAT GTTGATGTTT

      ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
  401 CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCCACC AACACCGCTG GCGCCAAGTA

      ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
  501 CGGCACGGGG TACTGTGACA GCCAGTGTCC CCGCGATCTG AAGTTCATCA ATGGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGCGAAC

      ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla PheThrPro HisProCys ThrThrValGly·
  601 ACGGGCATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTTTACCCC CCACCCTTGC ACGACTGTCG

      ·GGlnGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgTyrGlyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyrArg·
  701 GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGAACTTA CTCCGATAAC AGATATGGCG GCACTTGCGA TCCCGATGGC TGCGACTGGA ACCCATACCG

      ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaIle
  801 CCTGGGCAAC ACCAGCTTCT ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGCCATC

      AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
  901 AACCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

      ·GGluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp·
 1001 AGGAGGCAGA ATTCGGCGGA TCCTCTTTCT CAGACAAGGG CGGCCTGACT CAGTTCAAGA AGGCTACCTC TGGCGGCATG GTTCTGGTCA TGAGTCTGTG

      ·AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerThr TyrProThrAsn GluThrSer SerThrPro GlyAlaValArg GlySerCys SerThrSer
 1101 GGATGATTAC TACGCCAACA TGCTGTGGCT GGACTCCACC TACCCGACAA ACGAGACCTC CTCCACACCC GGTGCCGTGC GCGGAAGCTG CTCCACCAGC

      SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrGly AsnProSerGly·
 1201 TCCGGTGTCC CTGCTCAGGT CGAATCTCAG TCTCCCAACG CCAAGGTCAC CTTCTCCAAC ATCAAGTTCG GACCCATTGG CAGCACCGGC AACCCTAGCG

      ·GGlyAsnPro ProGlyGly AsnProProGly ThrThrThr ThrArgArg ProAlaThrThr ThrGlySer SerProGly ProThrGlnSer HisTyrGly·
 1301 GCGGCAACCC TCCCGGCGGA AACCCGCCTG GCACCACCAC CACCCGCCGC CCAGCCACTA CCACTGGAAG CTCTCCCGGA CCTACCCAGT CTCACTACGG

      ·GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaSerGlyThr ThrCysGln ValLeuAsn ProTyrTyrSer GlnCysLeu
 1401 CCAGTGCGGC GGTATTGGCT ACAGCGGCCC CACGGTCTGC GCCAGCGGCA CAACTTGCCA GGTCCTGAAC CCTTACTACT CTCAGTGCCT G
```

Figure 5B: AminoAcid and Nucleic Acid Sequence for *Hypocrea jecorina* Cel7A mutant S113D

```
      GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle·
   1  CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCCGCCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTTGCACTCA ACAGACAGGC TCCGTGGTCA

      ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
 101  TCGACGCCAA CTGGCGCTGG ACTCACGCTA CGAACAGCAG CACGAACTGC TACGATGGCA ACACTTGGAG CTCGACCCTA TGTCCTGACA ACGAGACCTG

      ·AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
 201  CGCGAAGAAC TGCTGTCTGG ACGGTGCCGC CTACGCGTCC ACGTACGGAG TTACCACGAG CGGTAACAGC CTCTCCATTG GCTTTGTCAC CCAGTCTGCG

      GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaAspAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
 301  CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGGACG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTTCGAT GTTGATGTTT

      ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
 401  CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCCACC AACACCGCTG GCGCCAAGTA

      ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
 501  CGGCACGGGG TACTGTGACA GCCAGTGTCC CCGCGATCTG AAGTTCATCA ATGGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGCGAAC

      ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly·
 601  ACGGGCATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTCTTACCCC CCACCCTTGC ACGACTGTCG

      ·GGlnGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgTyrGlyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyrArg·
 701  GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGAACTTA CTCCGATAAC AGATATGGCG GCACTTGCGA TCCCGATGGC TGCGACTGGA ACCCATACCG

      ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaIle
 801  CCTGGGCAAC ACCAGCTTCT ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGCCATC

      AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
 901  AACCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

      ·GGluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp·
1001  AGGAGGCAGA ATTCGGCGGA TCCTCTTTCT CAGACAAGGG CGGCCTGACT CAGTTCAAGA AGGCTACCTC TGGCGGCATG GTTCTGGTCA TGAGTCTGTG

      ·AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerThr TyrProThrAsn GluThrSer SerThrPro GlyAlaValArg GlySerCys SerThrSer
1101  GGATGATTAC TACGCCAACA TGCTGTGGCT GGACTCCACC TACCCGACAA ACGAGACCTC CTCCACACCC GGTGCCGTGC GCGGAAGCTG CTCCACCAGC

      SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrGly AsnProSerGly·
1201  TCCGGTGTCC CTGCTCAGGT CGAATCTCAG TCTCCCAACG CCAAGGTCAC CTTCTCCAAC ATCAAGTTCG GACCCATTGG CAGCACCGGC AACCCTAGCG

      ·GGlyAsnPro ProGlyGly AsnProProGly ThrThrThr ThrArgArg ProAlaThrThr ThrGlySer SerProGly ProThrGlnSer HisTyrGly·
1301  GCGGCAACCC TCCCGGCGGA AACCCGCCTG GCACCACCAC CACCCGCCGC CCAGCCACTA CCACTGGAAG CTCTCCCGGA CCTACCCAGT CTCACTACGG

      ·GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaSerGlyThr ThrCysGln ValLeuAsn ProTyrTyrSer GlnCysLeu
1401  CCAGTGCGGC GGTATTGGCT ACAGCGGCCC CACGGTCTGC GCCAGCGGCA CAACTTGCCA GGTCCTGAAC CCTTACTACT CTCAGTGCCT G
```

Figure 5C: AminoAcid and Nucleic Acid Sequence for *Hypocrea jecorina* Cel7A mutant A77D

```
      GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle·
   1  CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCCGCCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTTGCACTCA ACAGACAGGC TCCGTGGTCA

      ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
 101  TCGACGCCAA CTGGCGCTGG ACTCACGCTA CGAACAGCAG CACGAACTGC TACGATGGCA ACACTTGGAG CTCGACCCTA TGTCCTGACA ACGAGACCTG

      ·AlaLysAsn CysCysLeuAsp GlyAlaAsp TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
 201  CGCGAAGAAC TGCTGTCTGG ACGGTGCCGA CTACGCGTCC ACGTACGGAG TTACCACGAG CGGTAACAGC CTCTCCATTG GCTTTGTCAC CCAGTCTGCG

      GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
 301  CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGAGCG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTTCGAT GTTGATGTTT

      ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
 401  CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCCACC AACACCGCTG GCGCCAAGTA

      ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
 501  CGGCACGGGG TACTGTGACA GCCAGTGTCC CCGCGATCTG AAGTTCATCA ATGGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGCGAAC

      ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly·
 601  ACGGGCATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTCTTACCCC CCACCCTTGC ACGACTGTCG

      ·GGlnGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgTyrGlyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyrArg·
 701  GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGAACTTA CTCCGATAAC AGATATGGCG GCACTTGCGA TCCCGATGGC TGCGACTGGA ACCCATACCG

      ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaIle
 801  CCTGGGCAAC ACCAGCTTCT ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGCCATC

      AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
 901  AACCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

      ·GGluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp·
1001  AGGAGGCAGA ATTCGGCGGA TCCTCTTTCT CAGACAAGGG CGGCCTGACT CAGTTCAAGA AGGCTACCTC TGGCGGCATG GTTCTGGTCA TGAGTCTGTG

      ·AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerThr TyrProThrAsn GluThrSer SerThrPro GlyAlaValArg GlySerCys SerThrSer
1101  GGATGATTAC TACGCCAACA TGCTGTGGCT GGACTCCACC TACCCGACAA ACGAGACCTC CTCCACACCC GGTGCCGTGC GCGGAAGCTG CTCCACCAGC

      SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrGly AsnProSerGly·
1201  TCCGGTGTCC CTGCTCAGGT CGAATCTCAG TCTCCCAACG CCAAGGTCAC CTTCTCCAAC ATCAAGTTCG GACCCATTGG CAGCACCGGC AACCCTAGCG

      ·GGlyAsnPro ProGlyGly AsnProProGly ThrThrThr ThrArgArg ProAlaThrThr ThrGlySer SerProGly ProThrGlnSer HisTyrGly·
1301  GCGGCAACCC TCCCGGCGGA AACCCGCCTG GCACCACCAC CACCCGCCGC CCAGCCACTA CCACTGGAAG CTCTCCCGGA CCTACCCAGT CTCACTACGG

      ·GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaSerGlyThr ThrCysGln ValLeuAsn ProTyrTyrSer GlnCysLeu
1401  CCAGTGCGGC GGTATTGGCT ACAGCGGCCC CACGGTCTGC GCCAGCGGCA CAACTTGCCA GGTCCTGAAC CCTTACTACT CTCAGTGCCT G
```

Figure 5D: Amino Acid and Nucleic Acid Sequences of *Hypocrea jecorina* Cel7A mutant L288F

```
      GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle·
1     CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCCGCCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTTGCACTCA ACAGACAGGC TCCGTGGTCA

      ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
101   TCGACGCCAA CTGGCGCTGG ACTCACGCTA CGAACAGCAG CACGAACTGC TACGATGGCA ACACTTGGAG CTCGACCCTA TGTCCTGACA ACGAGACCTG

      ·AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
201   CGCGAAGAAC TGCTGTCTGG ACGGTGCCGC CTACGCGTCC ACGTACGGAG TTACCACGAG CGGTAACAGC CTCTCCATTG GCTTTGTCAC CCAGTCTGCG

      GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
301   CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGAGCG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTTCGAT GTTGATGTTT

      ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
401   CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCCACC AACACCGCTG GCGCCAAGTA

      ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
501   CGGCACGGGG TACTGTGACA GCCAGTGTCC CCGCGATCTG AAGTTCATCA ATGGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGCGAAC

      ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly·
601   ACGGGCATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTCTTACCCC CCACCCTTGC ACGACTGTCG

      ·GGlnGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgTyrGlyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyrArg·
701   GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGAACTTA CTCCGATAAC AGATATGGCG GCACTTGCGA TCCCGATGGC TGCGACTGGA ACCCATACCG

      ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys PheThrVal ValThrGlnPhe GluThrSer GlyAlaIle
801   CCTGGGCAAC ACCAGCTTCT ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTTACCGTT GTCACCCAGT TCGAGACGTC GGGTGCCATC

      AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
901   AACCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

      ·GGluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp·
1001  AGGAGGCAGA ATTCGGCGGA TCCTCTTTCT CAGACAAGGG CGGCCTGACT CAGTTCAAGA AGGCTACCTC TGGCGGCATG GTTCTGGTCA TGAGTCTGTG

      ·AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerThr TyrProThrAsn GluThrSer SerThrPro GlyAlaValArg GlySerCys SerThrSer
1101  GGATGATTAC TACGCCAACA TGCTGTGGCT GGACTCCACC TACCCGACAA ACGAGACCTC CTCCACACCC GGTGCCGTGC GCGGAAGCTG CTCCACCAGC

      SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrGly AsnProSerGly·
1201  TCCGGTGTCC CTGCTCAGGT CGAATCTCAG TCTCCCAACG CCAAGGTCAC CTTCTCCAAC ATCAAGTTCG GACCCATTGG CAGCACCGGC AACCCTAGCG

      ·GGlyAsnPro ProGlyGly AsnProProGly ThrThrThr ThrArgArg ProAlaThrThr ThrGlySer SerProGly ProThrGlnSer HisTyrGly·
1301  GCGGCAACCC TCCCGGCGGA AACCCGCCTG GCACCACCAC CACCCGCCGC CCAGCCACTA CCACTGGAAG CTCTCCCGGA CCTACCCAGT CTCACTACGG

      ·GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaSerGlyThr ThrCysGln ValLeuAsn ProTyrTyrSer GlnCysLeu
1401  CCAGTGCGGC GGTATTGGCT ACAGCGGCCC CACGGTCTGC GCCAGCGGCA CAACTTGCCA GGTCCTGAAC CCTTACTACT CTCAGTGCCT G
```

Figure 5E: AminoAcid and Nucleic Acid Sequence of *Hypocrea jecorina* Cel7A mutant A299E

```
      GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle·
1     CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCCGCCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTTGCACTCA ACAGACAGGC TCCGTGGTCA

      ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
101   TCGACGCCAA CTGGCGCTGG ACTCACGCTA CGAACAGCAG CACGAACTGC TACGATGGCA ACACTTGGAG CTCGACCCTA TGTCCTGACA ACGAGACCTG

      ·AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
201   CGCGAAGAAC TGCTGTCTGG ACGGTGCCGC CTACGCGTCC ACGTACGGAG TTACCACGAG CGGTAACAGC CTCTCCATTG GCTTTGTCAC CCAGTCTGCG

      GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
301   CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGAGCG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTTCGAT GTTGATGTTT

      ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
401   CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCCACC AACACCGCTG GCGCCAAGTA

      ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
501   CGGCACGGGG TACTGTGACA GCCAGTGTCC CCGCGATCTG AAGTTCATCA ATGGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGCGAAC

      ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly·
601   ACGGGCATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTCTTACCCC CCACCCTTGC ACGACTGTCG

      ·GGlnGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgTyrGlyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyrArg·
701   GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGAACTTA CTCCGATAAC AGATATGGCG GCACTTGCGA TCCCGATGGC TGCGACTGGA ACCCATACCG

      ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyGluIle
801   CCTGGGCAAC ACCAGCTTCT ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGAGATC

      AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
901   AACCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

      ·GGluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp·
1001  AGGAGGCAGA ATTCGGCGGA TCCTCTTTCT CAGACAAGGG CGGCCTGACT CAGTTCAAGA AGGCTACCTC TGGCGGCATG GTTCTGGTCA TGAGTCTGTG

      ·AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerThr TyrProThrAsn GluThrSer SerThrPro GlyAlaValArg GlySerCys SerThrSer
1101  GGATGATTAC TACGCCAACA TGCTGTGGCT GGACTCCACC TACCCGACAA ACGAGACCTC CTCCACACCC GGTGCCGTGC GCGGAAGCTG CTCCACCAGC

      SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrGly AsnProSerGly·
1201  TCCGGTGTCC CTGCTCAGGT CGAATCTCAG TCTCCCAACG CCAAGGTCAC CTTCTCCAAC ATCAAGTTCG GACCCATTGG CAGCACCGGC AACCCTAGCG

      ·GGlyAsnPro ProGlyGly AsnProProGly ThrThrThr ThrArgArg ProAlaThrThr ThrGlySer SerProGly ProThrGlnSer HisTyrGly·
1301  GCGGCAACCC TCCCGGCGGA AACCCGCCTG GCACCACCAC CACCCGCCGC CCAGCCACTA CCACTGGAAG CTCTCCCGGA CCTACCCAGT CTCACTACGG

      ·GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaSerGlyThr ThrCysGln ValLeuAsn ProTyrTyrSer GlnCysLeu
1401  CCAGTGCGGC GGTATTGGCT ACAGCGGCCC CACGGTCTGC GCCAGCGGCA CAACTTGCCA GGTCCTGAAC CCTTACTACT CTCAGTGCCT G
```

Figure 5F: AminoAcid and Nucleic Acid Sequence of *Hypocrea jecorina* Cel7A mutant N301K

```
      GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle·
    1 CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCCGCCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTTGCACTCA ACAGACAGGC TCCGTGGTCA

      ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
  101 TCGACGCCAA CTGGCGCTGG ACTCACGCTA CGAACAGCAG CACGAACTGC TACGATGGCA ACACTTGGAG CTCGACCCTA TGTCCTGACA ACGAGACCTG

      ·AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
  201 CGCGAAGAAC TGCTGTCTGG ACGGTGCCGC CTACGCGTCC ACGTACGGAG TTACCACGAG CGGTAACAGC CTCTCCATTG GCTTTGTCAC CCAGTCTGCG

      GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
  301 CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGAGCG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTTCGAT GTTGATGTTT

      ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
  401 CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCCACC AACACCGCTG GCGCCAAGTA

      ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
  501 CGGCACGGGG TACTGTGACA GCCAGTGTCC CCGCGATCTG AAGTTCATCA ATGGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGCGAAC

      ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly·
  601 ACGGGCATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTCTTACCCC CCACCCTTGC ACGACTGTCG

      ·GGlnGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgTyrGlyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyrArg·
  701 GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGAACTTA CTCCGATAAC AGATATGGCG GCACTTGCGA TCCCGATGGC TGCGACTGGA ACCCATACCG

      ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaIle
  801 CCTGGGCAAC ACCAGCTTCT ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGCCATC

      LysArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
  901 AAGCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

      ·GGluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp·
 1001 AGGAGGCAGA ATTCGGCGGA TCCTCTTTCT CAGACAAGGG CGGCCTGACT CAGTTCAAGA AGGCTACCTC TGGCGGCATG GTTCTGGTCA TGAGTCTGTG

      ·AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerThr TyrProThrAsn GluThrSer SerThrPro GlyAlaValArg GlySerCys SerThrSer
 1101 GGATGATTAC TACGCCAACA TGCTGTGGCT GGACTCCACC TACCCGACAA ACGAGACCTC CTCCACACCC GGTGCCGTGC GCGGAAGCTG CTCCACCAGC

      SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrGly AsnProSerGly·
 1201 TCCGGTGTCC CTGCTCAGGT CGAATCTCAG TCTCCCAACG CCAAGGTCAC CTTCTCCAAC ATCAAGTTCG GACCCATTGG CAGCACCGGC AACCCTAGCG

      ·GGlyAsnPro ProGlyGly AsnProProGly ThrThrThr ThrArgArg ProAlaThrThr ThrGlySer SerProGly ProThrGlnSer HisTyrGly·
 1301 GCGGCAACCC TCCCGGCGGA AACCCGCCTG GCACCACCAC CACCCGCCGC CCAGCCACTA CCACTGGAAG CTCTCCCGGA CCTACCCAGT CTCACTACGG

      ·GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaSerGlyThr ThrCysGln ValLeuAsn ProTyrTyrSer GlnCysLeu
 1401 CCAGTGCGGC GGTATTGGCT ACAGCGGCCC CACGGTCTGC GCCAGCGGCA CAACTTGCCA GGTCCTGAAC CCTTACTACT CTCAGTGCCT G
```

Figure 5G: AminoAcid and Nucleic Acid Sequence of *Hypocrea jecorina* Cel7A mutant T356L

```
      GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle·
1     CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCCGCCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTTGCACTCA ACAGACAGGC TCCGTGGTCA

      ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
101   TCGACGCCAA CTGGCGCTGG ACTCACGCTA CGAACAGCAG CACGAACTGC TACGATGGCA ACACTTGGAG CTCGACCCTA TGTCCTGACA ACGAGACCTG

      ·AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
201   CGCGAAGAAC TGCTGTCTGG ACGGTGCCGC CTACGCGTCC ACGTACGGAG TTACCACGAG CGGTAACAGC CTCTCCATTG GCTTTGTCAC CCAGTCTGCG

      GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
301   CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGAGCG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTTCGAT GTTGATGTTT

      ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
401   CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCCACC AACACCGCTG GCGCCAAGTA

      ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
501   CGGCACGGGG TACTGTGACA GCCAGTGTCC CCGCGATCTG AAGTTCATCA ATGGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGCGAAC

      ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly·
601   ACGGGCATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTCTTACCCC CCACCCTTGC ACGACTGTCG

      ·GGlnGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgTyrGlyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyrArg·
701   GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGAACTTA CTCCGATAAC AGATATGGCG GCACTTGCGA TCCCGATGGC TGCGACTGGA ACCCATACCG

      ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaIle
801   CCTGGGCAAC ACCAGCTTCT ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGCCATC

      AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
901   AACCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

      ·GGluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaLeuSer GlyGlyMet ValLeuValMet SerLeuTrp·
1001  AGGAGGCAGA ATTCGGCGGA TCCTCTTTCT CAGACAAGGG CGGCCTGACT CAGTTCAAGA AGGCTCTCTC TGGCGGCATG GTTCTGGTCA TGAGTCTGTG

      ·AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerThr TyrProThrAsn GluThrSer SerThrPro GlyAlaValArg GlySerCys SerThrSer
1101  GGATGATTAC TACGCCAACA TGCTGTGGCT GGACTCCACC TACCCGACAA ACGAGACCTC CTCCACACCC GGTGCCGTGC GCGGAAGCTG CTCCACCAGC

      SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrGly AsnProSerGly·
1201  TCCGGTGTCC CTGCTCAGGT CGAATCTCAG TCTCCCAACG CCAAGGTCAC CTTCTCCAAC ATCAAGTTCG GACCCATTGG CAGCACCGGC AACCCTAGCG

      ·GGlyAsnPro ProGlyGly AsnProProGly ThrThrThr ThrArgArg ProAlaThrThr ThrGlySer SerProGly ProThrGlnSer HisTyrGly·
1301  GCGGCAACCC TCCCGGCGGA AACCCGCCTG GCACCACCAC CACCCGCCGC CCAGCCACTA CCACTGGAAG CTCTCCCGGA CCTACCCAGT CTCACTACGG

      ·GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaSerGlyThr ThrCysGln ValLeuAsn ProTyrTyrSer GlnCysLeu
1401  CCAGTGCGGC GGTATTGGCT ACAGCGGCCC CACGGTCTGC GCCAGCGGCA CAACTTGCCA GGTCCTGAAC CCTTACTACT CTCAGTGCCT G
```

Figure 5H: AminoAcid and Nucleic Acid Sequence of *Hypocrea jecorina* Cel7A mutant G430F

```
      GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle·
1     CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCCGCCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTTGCACTCA ACAGACAGGC TCCGTGGTCA

      ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
101   TCGACGCCAA CTGGCGCTGG ACTCACGCTA CGAACAGCAG CACGAACTGC TACGATGGCA ACACTTGGAG CTCGACCCTA TGTCCTGACA ACGAGACCTG

      ·AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
201   CGCGAAGAAC TGCTGTCTGG ACGGTGCCGC CTACGCGTCC ACGTACGGAG TTACCACGAG CGGTAACAGC CTCTCCATTG GCTTTGTCAC CCAGTCTGCG

      GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
301   CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGAGCG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTTCGAT GTTGATGTTT

      ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
401   CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCCACC AACACCGCTG GCGCCAAGTA

      ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
501   CGGCACGGGG TACTGTGACA GCCAGTGTCC CCGCGATCTG AAGTTCATCA ATGGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGCGAAC

      ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly·
601   ACGGGCATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTCTTACCCC CCACCCTTGC ACGACTGTCG

      ·GGlnGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgTyrGlyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyrArg·
701   GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGAACTTA CTCCGATAAC AGATATGGCG GCACTTGCGA TCCCGATGGC TGCGACTGGA ACCCATACCG

      ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaIle
801   CCTGGGCAAC ACCAGCTTCT ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGCCATC

      AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
901   AACCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

      ·GGluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp·
1001  AGGAGGCAGA ATTCGGCGGA TCCTCTTTCT CAGACAAGGG CGGCCTGACT CAGTTCAAGA AGGCTACCTC TGGCGGCATG GTTCTGGTCA TGAGTCTGTG

      ·AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerThr TyrProThrAsn GluThrSer SerThrPro GlyAlaValArg GlySerCys SerThrSer
1101  GGATGATTAC TACGCCAACA TGCTGTGGCT GGACTCCACC TACCCGACAA ACGAGACCTC CTCCACACCC GGTGCCGTGC GCGGAAGCTG CTCCACCAGC

      SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrPhe AsnProSerGly·
1201  TCCGGTGTCC CTGCTCAGGT CGAATCTCAG TCTCCCAACG CCAAGGTCAC CTTCTCCAAC ATCAAGTTCG GACCCATTGG CAGCACCTTC AACCCTAGCG

      ·GGlyAsnPro ProGlyGly AsnProProGly ThrThrThr ThrArgArg ProAlaThrThr ThrGlySer SerProGly ProThrGlnSer HisTyrGly·
1301  GCGGCAACCC TCCCGGCGGA AACCCGCCTG GCACCACCAC CACCCGCCGC CCAGCCACTA CCACTGGAAG CTCTCCCGGA CCTACCCAGT CTCACTACGG

      ·GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaSerGlyThr ThrCysGln ValLeuAsn ProTyrTyrSer GlnCysLeu
1401  CCAGTGCGGC GGTATTGGCT ACAGCGGCCC CACGGTCTGC GCCAGCGGCA CAACTTGCCA GGTCCTGAAC CCTTACTACT CTCAGTGCCT G
```

Figure 5l: AminoAcid and Nucleic Acid Sequence for *Hypocrea jecorina* Cel7A mutant T246C/Y371C

```
                GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle·
1               CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCCGCCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTTGCACTCA ACAGACAGGC TCCGTGGTCA

                ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
101             TCGACGCCAA CTGGCGCTGG ACTCACGCTA CGAACAGCAG CACGAACTGC TACGATGGCA ACACTTGGAG CTCGACCCTA TGTCCTGACA ACGAGACCTG

                ·AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
201             CGCGAAGAAC TGCTGTCTGG ACGGTGCCGC CTACGCGTCC ACGTACGGAG TTACCACGAG CGGTAACAGC CTCTCCATTG GCTTTGTCAC CCAGTCTGCG

                GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
301             CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGAGCG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTTCGAT GTTGATGTTT

                ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
401             CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCCACC AACACCGCTG GCGCCAAGTA

                ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
501             CGGCACGGGG TACTGTGACA GCCAGTGTCC CCGCGATCTG AAGTTCATCA ATGGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGCGAAC

                ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly·
601             ACGGGCATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTCTTACCCC CCACCCTTGC ACGACTGTCG

                ·GGlnGluIle CysGluGly AspGlyCysGly GlyCysTyr SerAspAsn ArgTyrGlyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyrArg·
701             GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGATGTTA CTCCGATAAC AGATATGGCG GCACTTGCGA TCCCGATGGC TGCGACTGGA ACCCATACCG

                ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaIle
801             CCTGGGCAAC ACCAGCTTCT ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGCCATC

                AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
901             AACCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

                ·GGluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp·
1001            AGGAGGCAGA ATTCGGCGGA TCCTCTTTCT CAGACAAGGG CGGCCTGACT CAGTTCAAGA AGGCTACCTC TGGCGGCATG GTTCTGGTCA TGAGTCTGTG

                ·AspAspTyr CysAlaAsnMet LeuTrpLeu AspSerThr TyrProThrAsn GluThrSer SerThrPro GlyAlaValArg GlySerCys SerThrSer
1101            GGATGATTAC TGCGCCAACA TGCTGTGGCT GGACTCCACC TACCCGACAA ACGAGACCTC CTCCACACCC GGTGCCGTGC GCGGAAGCTG CTCCACCAGC

                SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrGly AsnProSerGly·
1201            TCCGGTGTCC CTGCTCAGGT CGAATCTCAG TCTCCCAACG CCAAGGTCAC CTTCTCCAAC ATCAAGTTCG GACCCATTGG CAGCACCGGC AACCCTAGCG

                ·GGlyAsnPro ProGlyGly AsnProProGly ThrThrThr ThrArgArg ProAlaThrThr ThrGlySer SerProGly ProThrGlnSer HisTyrGly·
1301            GCGGCAACCC TCCCGGCGGA AACCCGCCTG GCACCACCAC CACCCGCCGC CCAGCCACTA CCACTGGAAG CTCTCCCGGA CCTACCCAGT CTCACTACGG

                ·GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaSerGlyThr ThrCysGln ValLeuAsn ProTyrTyrSer GlnCysLeu
1401            CCAGTGCGGC GGTATTGGCT ACAGCGGCCC CACGGTCTGC GCCAGCGGCA CAACTTGCCA GGTCCTGAAC CCTTACTACT CTCAGTGCCT G
```

Figure 5J: AminoAcid and Nucleic Acid Sequence for *Hypocrea jecorina* Cel7A mutant T246A/R251A/Y252A

```
      GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle·
1     CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCCGCCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTTGCACTCA ACAGACAGGC TCCGTGGTCA

      ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
101   TCGACGCCAA CTGGCGCTGG ACTCACGCTA CGAACAGCAG CACGAACTGC TACGATGGCA ACACTTGGAG CTCGACCCTA TGTCCTGACA ACGAGACCTG

      ·AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
201   CGCGAAGAAC TGCTGTCTGG ACGGTGCCGC CTACGCGTCC ACGTACGGAG TTACCACGAG CGGTAACAGC CTCTCCATTG GCTTTGTCAC CCAGTCTGCG

      GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
301   CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGAGCG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTTCGAT GTTGATGTTT

      ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
401   CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCCACC AACACCGCTG GCGCCAAGTA

      ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
501   CGGCACGGGG TACTGTGACA GCCAGTGTCC CCGCGATCTG AAGTTCATCA ATGGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGCGAAC

      ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly·
601   ACGGGCATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTCTTACCCC CCACCCTTGC ACGACTGTCG

      ·GGlnGluIle CysGluGly AspGlyCysGly GlyAlaTyr SerAspAsn AlaAlaGlyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyrArg·
701   GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGAGCTTA CTCCGATAAC GCAGCTGGCG GCACTTGCGA TCCCGATGGC TGCGACTGGA ACCCATACCG

      ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaIle
801   CCTGGGCAAC ACCAGCTTCT ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGCCATC

      AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
901   AACCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

      ·GGluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp·
1001  AGGAGGCAGA ATTCGGCGGA TCCTCTTTCT CAGACAAGGG CGGCCTGACT CAGTTCAAGA AGGCTACCTC TGGCGGCATG GTTCTGGTCA TGAGTCTGTG

      ·AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerThr TyrProThrAsn GluThrSer SerThrPro GlyAlaValArg GlySerCys SerThrSer
1101  GGATGATTAC TACGCCAACA TGCTGTGGCT GGACTCCACC TACCCGACAA ACGAGACCTC CTCCACACCC GGTGCCGTGC GCGGAAGCTG CTCCACCAGC

      SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrGly AsnProSerGly·
1201  TCCGGTGTCC CTGCTCAGGT CGAATCTCAG TCTCCCAACG CCAAGGTCAC CTTCTCCAAC ATCAAGTTCG GACCCATTGG CAGCACCGGC AACCCTAGCG

      ·GGlyAsnPro ProGlyGly AsnProProGly ThrThrThr ThrArgArg ProAlaThrThr ThrGlySer SerProGly ProThrGlnSer HisTyrGly·
1301  GCGGCAACCC TCCCGGCGGA AACCCGCCTG GCACCACCAC CACCCGCCGC CCAGCCACTA CCACTGGAAG CTCTCCCGGA CCTACCCAGT CTCACTACGG

      ·GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaSerGlyThr ThrCysGln ValLeuAsn ProTyrTyrSer GlnCysLeu
1401  CCAGTGCGGC GGTATTGGCT ACAGCGGCCC CACGGTCTGC GCCAGCGGCA CAACTTGCCA GGTCCTGAAC CCTTACTACT CTCAGTGCCT G
```

Figure 5K: AminoAcid and Nucleic Acid Sequence of *Hypocrea jecorina* Cel7A mutant T380G/Y381D/R394A

```
      GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle·
1     CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCCGCCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTTGCACTCA ACAGACAGGC TCCGTGGTCA

      ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
101   TCGACGCCAA CTGGCGCTGG ACTCACGCTA CGAACAGCAG CACGAACTGC TACGATGGCA ACACTTGGAG CTCGACCCTA TGTCCTGACA ACGAGACCTG

      ·AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
201   CGCGAAGAAC TGCTGTCTGG ACGGTGCCGC CTACGCGTCC ACGTACGGAG TTACCACGAG CGGTAACAGC CTCTCCATTG GCTTTGTCAC CCAGTCTGCG

      GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
301   CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGAGCG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTTCGAT GTTGATGTTT

      ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
401   CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCCACC AACACCGCTG GCGCCAAGTA

      ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
501   CGGCACGGGG TACTGTGACA GCCAGTGTCC CCGCGATCTG AAGTTCATCA ATGGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGCGAAC

      ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly·
601   ACGGGCATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTCTTACCCC CCACCCTTGC ACGACTGTCG

      ·GGlnGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgTyrGlyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyrArg·
701   GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGAACTTA CTCCGATAAC AGATATGGCG GCACTTGCGA TCCCGATGGC TGCGACTGGA ACCCATACCG

      ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaIle
801   CCTGGGCAAC ACCAGCTTCT ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGCCATC

      AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
901   AACCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

      ·GGluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp·
1001  AGGAGGCAGA ATTCGGCGGA TCCTCTTTCT CAGACAAGGG CGGCCTGACT CAGTTCAAGA AGGCTACCTC TGGCGGCATG GTTCTGGTCA TGAGTCTGTG

      ·AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerGly AspProThrAsn GluThrSer SerThrPro GlyAlaValAla GlySerCys SerThrSer
101   GGATGATTAC TACGCCAACA TGCTGTGGCT GGACTCCGGC GACCCGACAA ACGAGACCTC CTCCACACCC GGTGCCGTGG CCGGAAGCTG CTCCACCAGC

      SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrGly AsnProSerGly·
1201  TCCGGTGTCC CTGCTCAGGT CGAATCTCAG TCTCCCAACG CCAAGGTCAC CTTCTCCAAC ATCAAGTTCG GACCCATTGG CAGCACCGGC AACCCTAGCG

      ·GGlyAsnPro ProGlyGly AsnProProGly ThrThrThr ThrArgArg ProAlaThrThr ThrGlySer SerProGly ProThrGlnSer HisTyrGly·
1301  GCGGCAACCC TCCCGGCGGA AACCCGCCTG GCACCACCAC CACCCGCCGC CCAGCCACTA CCACTGGAAG CTCTCCCGGA CCTACCCAGT CTCACTACGG

      ·GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaSerGlyThr ThrCysGln ValLeuAsn ProTyrTyrSer GlnCysLeu
1401  CCAGTGCGGC GGTATTGGCT ACAGCGGCCC CACGGTCTGC GCCAGCGGCA CAACTTGCCA GGTCCTGAAC CCTTACTACT CTCAGTGCCT G
```

Figure 5L: AminoAcid and Nucleic Acid Sequence of *Hypocrea jecorina* Cel7A mutant T380G/Y381D/R394A with residues 381 through 393 deleted

```
                GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle·
1               CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCCGCCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTTGCACTCA ACAGACAGGC TCCGTGGTCA

                ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
101             TCGACGCCAA CTGGCGCTGG ACTCACGCTA CGAACAGCAG CACGAACTGC TACGATGGCA ACACTTGGAG CTCGACCCTA TGTCCTGACA ACGAGACCTG

                ·AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
201             CGCGAAGAAC TGCTGTCTGG ACGGTGCCGC CTACGCGTCC ACGTACGGAG TTACCACGAG CGGTAACAGC CTCTCCATTG GCTTTGTCAC CCAGTCTGCG

                GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
301             CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGAGCG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTTCGAT GTTGATGTTT

                ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
401             CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCCACC AACACCGCTG GCGCCAAGTA

                ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
501             CGGCACGGGG TACTGTGACA GCCAGTGTCC CCGCGATCTG AAGTTCATCA ATGGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGCGAAC

                ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly·
601             ACGGGCATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTCTTACCCC CCACCCTTGC ACGACTGTCG

                ·GGlnGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgTyrGlyGly ThrCysAsp ProAspGly CysAspTrpAsn ProTyrArg·
701             GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGAACTTA CTCCGATAAC AGATATGGCG GCACTTGCGA TCCCGATGGC TGCGACTGGA ACCCATACCG

                ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaIle
801             CCTGGGCAAC ACCAGCTTCT ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGCCATC

                AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
901             AACCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

                ·GGluAlaGlu PheGlyGly SerSerPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp·
1001            AGGAGGCAGA ATTCGGCGGA TCCTCTTTCT CAGACAAGGG CGGCCTGACT CAGTTCAAGA AGGCTACCTC TGGCGGCATG GTTCTGGTCA TGAGTCTGTG

                ·AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerGly AspAlaGlySer CysSerThr SerSerGly ValProAlaGln ValGluSer GlnSerPro
1101            GGATGATTAC TACGCCAACA TGCTGTGGCT GGACTCCGGC GACGCCGGAA GCTGCTCCAC CAGCTCCGGT GTCCCTGCTC AGGTCGAATC TCAGTCTCCC

                AsnAlaLysVal ThrPheSer AsnIleLys PheGlyProIle GlySerThr GlyAsnPro SerGlyGlyAsn ProProGly GlyAsnPro ProGlyThrThr·
1201            AACGCCAAGG TCACCTTCTC CAACATCAAG TTCGGACCCA TTGGCAGCAC CGGCAACCCT AGCGGCGGCA ACCCTCCCGG CGGAAACCCG CCTGGCACCA

                ·TThrThrArg ArgProAla ThrThrThrGly SerSerPro GlyProThr GlnSerHisTyr GlyGlnCys GlyGlyIle GlyTyrSerGly ProThrVal·
1301            CCACCACCCG CCGCCCAGCC ACTACCACTG GAAGCTCTCC CGGACCTACC CAGTCTCACT ACGGCCAGTG CGGCGGTATT GGCTACAGCG GCCCCACGGT

                ·CysAlaSer GlyThrThrCys GlnValLeu AsnProTyr TyrSerGlnCys Leu
1401            CTGCGCCAGC GGCACAACTT GCCAGGTCCT GAACCCTTAC TACTCTCAGT GCCTG
```

Figure 5M: AminoAcid and Nucleic Acid Sequence of *Hypocrea jecorina* Cel7A mutant Y252Q/D259W/S342Y

```
      GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValIle·
   1  CAGTCGGCCT GCACTCTCCA ATCGGAGACT CACCCGCCTC TGACATGGCA GAAATGCTCG TCTGGTGGCA CTTGCACTCA ACAGACAGGC TCCGTGGTCA

      ·IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys·
 101  TCGACGCCAA CTGGCGCTGG ACTCACGCTA CGAACAGCAG CACGAACTGC TACGATGGCA ACACTTGGAG CTCGACCCTA TGTCCTGACA ACGAGACCTG

      ·AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrGlyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla
 201  CGCGAAGAAC TGCTGTCTGG ACGGTGCCGC CTACGCGTCC ACGTACGGAG TTACCACGAG CGGTAACAGC CTCTCCATTG GCTTTGTCAC CCAGTCTGCG

      GlnLysAsnVal GlyAlaArg LeuTyrLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer·
 301  CAGAAGAACG TTGGCGCTCG CCTTTACCTT ATGGCGAGCG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTTCGAT GTTGATGTTT

      ·SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr·
 401  CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCCACC AACACCGCTG GCGCCAAGTA

      ·GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn
 501  CGGCACGGGG TACTGTGACA GCCAGTGTCC CCGCGATCTG AAGTTCATCA ATGGCCAGGC CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGCGAAC

      ThrGlyIleGly GlyHisGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly·
 601  ACGGGCATTG GAGGACACGG AAGCTGCTGC TCTGAGATGG ATATCTGGGA GGCCAACTCC ATCTCCGAGG CTCTTACCCC CCACCCTTGC ACGACTGTCG

      ··GGlnGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgGlnGlyGly ThrCysAsp ProTrpGly CysAspTrpAsn ProTyrArg·
 701  GCCAGGAGAT CTGCGAGGGT GATGGGTGCG GCGGAACTTA CTCCGATAAC AGACAGGGCG GCACTTGCGA TCCCTGGGGC TGCGACTGGA ACCCATACCG

      ·LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaIle
 801  CCTGGGCAAC ACCAGCTTCT ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGCCATC

      AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu·
 901  AACCGATACT ATGTCCAGAA TGGCGTCACT TTCCAGCAGC CCAACGCCGA GCTTGGTAGT TACTCTGGCA ACGAGCTCAA CGATGATTAC TGCACAGCTG

      ·GGluAlaGlu PheGlyGly SerTyrPheSer AspLysGly GlyLeuThr GlnPheLysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp·
1001  AGGAGGCAGA ATTCGGCGGA TCCTATTTCT CAGACAAGGG CGGCCTGACT CAGTTCAAGA AGGCTACCTC TGGCGGCATG GTTCTGGTCA TGAGTCTGTG

      ·AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerThr TyrProThrAsn GluThrSer SerThrPro GlyAlaValArg GlySerCys SerThrSer
1101  GGATGATTAC TACGCCAACA TGCTGTGGCT GGACTCCACC TACCCGACAA ACGAGACCTC CTCCACACCC GGTGCCGTGC GCGGAAGCTG CTCCACCAGC

      SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProIleGly SerThrGly AsnProSerGly·
1201  TCCGGTGTCC CTGCTCAGGT CGAATCTCAG TCTCCCAACG CCAAGGTCAC CTTCTCCAAC ATCAAGTTCG GACCCATTGG CAGCACCGGC AACCCTAGCG

      ·GGlyAsnPro ProGlyGly AsnProProGly ThrThrThr ThrArgArg ProAlaThrThr ThrGlySer SerProGly ProThrGlnSer HisTyrGly·
1301  GCGGCAACCC TCCCGGCGGA AACCCGCCTG GCACCACCAC CACCCGCCGC CCAGCCACTA CCACTGGAAG CTCTCCCGGA CCTACCCAGT CTCACTACGG

      ·GlnCysGly GlyIleGlyTyr SerGlyPro ThrValCys AlaSerGlyThr ThrCysGln ValLeuAsn ProTyrTyrSer GlnCysLeu
1401  CCAGTGCGGC GGTATTGGCT ACAGCGGCCC CACGGTCTGC GCCAGCGGCA CAACTTGCCA GGTCCTGAAC CCTTACTACT CTCAGTGCCT G
```

FIGURE 6A: pTEX2

```
AAGCTTAAGG TGCACGGCCC ACGTGGCCAC TAGTACTTCT CGAGCTCTGT      50
ACATGTCCGG TCGCGACGTA CGCGTATCGA TGGCGCCAGC TGCAGGCGGC     100
CGCCTGCAGC CACTTGCAGT CCCGTGGAAT TCTCACGGTG AATGTAGGCC     150
TTTTGTAGGG TAGGAATTGT CACTCAAGCA CCCCCAACCT CCATTACGCC     200
TCCCCCATAG AGTTCCCAAT CAGTGAGTCA TGGCACTGTT CTCAAATAGA     250
TTGGGGAGAA GTTGACTTCC GCCCAGAGCT GAAGGTCGCA CAACCGCATG     300
ATATAGGGTC GGCAACGGCA AAAAAGCACG TGGCTCACCG AAAAGCAAGA     350
TGTTTGCGAT CTAACATCCA GGAACCTGGA TACATCCATC ATCACGCACG     400
ACCACTTTGA TCTGCTGGTA AACTCGTATT CGCCCTAAAC CGAAGTGCGT     450
GGTAAATCTA CACGTGGGCC CCTTTCGGTA TACTGCGTGT GTCTTCTCTA     500
GGTGCCATTC TTTTCCCTTC CTCTAGTGTT GAATTGTTTG TGTTGGAGTC     550
CGAGCTGTAA CTACCTCTGA ATCTCTGGAG AATGGTGGAC TAACGACTAC     600
CGTGCACCTG CATCATGTAT ATAATAGTGA TCCTGAGAAG GGGGGTTTGG     650
AGCAATGTGG GACTTTGATG GTCATCAAAC AAAGAACGAA GACGCCTCTT     700
TTGCAAAGTT TTGTTTCGGC TACGGTGAAG AACTGGATAC TTGTTGTGTC     750
TTCTGTGTAT TTTTGTGGCA ACAAGAGGCC AGAGACAATC TATTCAAACA     800
CCAAGCTTGC TCTTTTGAGC TACAAGAACC TGTGGGGTAT ATATCTAGAG     850
TTGTGAAGTC GGTAATCCCG CTGTATAGTA ATACGAGTCG CATCTAAATA     900
CTCCGAAGCT GCTGCGAACC CGGAGAATCG AGATGTGCTG GAAAGCTTCT     950
AGCGAGCGGC TAAATTAGCA TGAAAGGCTA TGAGAAATTC TGGAGACGGC    1000
TTGTTGAATC ATGGCGTTCC ATTCTTCGAC AAGCAAAGCG TTCCGTCGCA    1050
GTAGCAGGCA CTCATTCCCG AAAAAACTCG GAGATTCCTA AGTAGCGATG    1100
GAACCGGAAT AATATAATAG GCAATACATT GAGTTGCCTC GACGGTTGCA    1150
ATGCAGGGGT ACTGAGCTTG GACATAACTG TTCCGTACCC CACCTCTTCT    1200
CAACCTTTGG CGTTTCCCTG ATTCAGCGTA CCCGTACAAG TCGTAATCAC    1250
TATTAACCCA GACTGACCGG ACGTGTTTTG CCCTTCATTT GGAGAAATAA    1300
TGTCATTGCG ATGTGTAATT TGCCTGCTTG ACCGACTGGG GCTGTTCGAA    1350
GCCCGAATGT AGGATTGTTA TCCGAACTCT GCTCGTAGAG GCATGTTGTG    1400
AATCTGTGTC GGGCAGGACA CGCCTCGAAG GTTCACGGCA AGGGAAACCA    1450
CCGATAGCAG TGTCTAGTAG CAACCTGTAA AGCCGCAATG CAGCATCACT    1500
GGAAAATACA AACCAATGGC TAAAAGTACA TAAGTTAATG CCTAAAGAAG    1550
TCATATACCA GCGGCTAATA ATTGTACAAT CAAGTGGCTA AACGTACCGT    1600
AATTTGCCAA CGGCTTGTGG GGTTGCAGAA GCAACGGCAA AGCCCCACTT    1650
CCCCACGTTT GTTTCTTCAC TCAGTCCAAT CTCAGCTGGT GATCCCCCAA    1700
TTGGGTCGCT TGTTTGTTCC GGTGAAGTGA AAGAAGACAG AGGTAAGAAT    1750
GTCTGACTCG GAGCGTTTTG CATACAACCA AGGGCAGTGA TGGAAGACAG    1800
TGAAATGTTG ACATTCAAGG AGTATTTAGC CAGGGATGCT TGAGTGTATC    1850
GTGTAAGGAG GTTTGTCTGC CGATACGACG AATACTGTAT AGTCACTTCT    1900
GATGAAGTGG TCCATATTGA AATGTAAGTC GGCACTGAAC AGGCAAAAGA    1950
TTGAGTTGAA ACTGCCTAAG ATCTCGGGCC CTCGGGCCTT CGGCCTTTGG    2000
GTGTACATGT TTGTGCTCCG GGCAAATGCA AAGTGTGGTA GGATCGAACA    2050
CACTGCTGCC TTTACCAAGC AGCTGAGGGT ATGTGATAGG CAAATGTTCA    2100
GGGGCCACTG CATGGTTTCG AATAGAAAGA GAAGCTTAGC CAAGAACAAT    2150
AGCCGATAAA GATAGCCTCA TTAAACGGAA TGAGCTAGTA GGCAAAGTCA    2200
GCGAATGTGT ATATATAAAG GTTCGAGGTC CGTGCCTCCC TCATGCTCTC    2250
CCCATCTACT CATCAACTCA GATCCTCCAG GAGACTTGTA CACCATCTTT    2300
TGAGGCACAG AAACCCAATA GTCAACCGCG GTTTAGGCGC GCCAGCTCCG    2350
TGCGAAAGCC TGACGCACCG GTAGATTCTT GGTGAGCCCG TATCATGACG    2400
GCGGCGGGAG CTACATGGCC CCGGGTGATT TATTTTTTTT GTATCTACTT    2450
```

FIGURE 6B: pTEX2

CTGACCCTTT TCAAATATAC GGTCAACTCA TCTTTCACTG GAGATGCGGC 2500
CTGCTTGGTA TTGCGATGTT GTCAGCTTGG CAAATTGTGG CTTTCGAAAA 2550
CACAAAACGA TTCCTTAGTA GCCATGCATT TTAAGATAAC GGAATAGAAG 2600
AAAGAGGAAA TTAAAAAAAA AAAAAAAACA AACATCCCGT TCATAACCCG 2650
TAGAATCGCC GCTCTTCGTG TATCCCAGTA CCAGTTTAAA CGGATCTCAA 2700
GCTTGCATGC AAAGATACAC ATCAATCGCA GCTGGGGTAC AATCATCCAT 2750
CATCCCAACT GGTACGTCAT AACAAAAATC GACAAGATGG AAAAAGAGGT 2800
CGCCTAAATA CAGCTGCATT CTATGATGCC GGGCTTTGGA CAAGAGCTCT 2850
TTCTCAGCTC CGTTTGTCCT CCCTCCCTTT TCCCCCTTCT TGCTAAATGC 2900
CTTTCTTTAC TTCTTTCTTC CCTTCCCTCC CCTATCGCAG CAGCCTCTCG 2950
GTGTAGGCTT TCCACGCTGC TGATCGGTAC CGCTCTGCCT CCTCTACGGG 3000
GTCTGAGGCC TTGAGGATGC CCCGGCCCAC AATGGCAATG TCGCTGCCGG 3050
CGATGCCAAT CAGCTTGTGC GGCGTGTTGT ACTGCTGGCC CTGGCCGTCT 3100
CCACCGACCG ATCCGTTGGT CTGCTGGTCC TCGTCTTCGG GGGGCAGCTG 3150
GCAGCCGGGC GTCATGTGGA TAAAGGCATC GTCGGGCTCG GTGTTGAGCG 3200
TCTCCTGCGA GATGAAGCCC ATGACAAAGT CCTTGTGCTC CCGGGCGGCC 3250
TCGACGCAGG CCTGCGTGTA CTCCTTGTTC ATGAAGTTGC CCTGGCTGGA 3300
CATTTGGGCG AGGATCAGGA GGCCTCGGCT CAGCGGCGCC TCCTCGATGC 3350
CCGGGAAGAG CGACTCGTCG CCCTCGGCGA TGGCCTTTGT TAACCGGGGC 3400
GAGGAGACGG ACTCGTACTG CTGGGTGACG GTGGTGATGG AGACGATGCT 3450
GCCCTTGCGG CCGTCGCCGG ACCGGTTCGA GTAGATGGGC TTGTCCAGGA 3500
CGCCAATGGA GCCCATGCCG TTGACGGCGC CGGCGGGCTC GGCGTCCCTG 3550
GAGTCGGCGT CGTCGTCAAA CGAGTCCATG GTGGGCGTGC CGACGGTGAC 3600
GGACGTCTTG ACCTCGCAGG GGTAGCGCTC GAGCCAGCGC TTGGCGCCCT 3650
GGGCCAGCGA GGCCACCGAC GCCTTGCCGG GCACCATGTT GACGTTGACA 3700
ATGTGCGCCC AGTCGATGAT GCGCGCCGAC CCGCCCGTGT ACTGCAGCTC 3750
GACGGTGTGG CCAATGTCGC CAAACTTGCG GTCCTCGAAG ATGAGGAAGC 3800
CGTGCTTGCG CGCCAGCGAC GCCAGCTGGG CTCCCGTGCC CGTCTCCGGG 3850
TGGAAGTCCC AGCCCGAGAC CATGTCGTAG TGCGTCTTGA GCACGACAAT 3900
CGACGGGCCA ATCTTGTCGG CCAGGTACAG CAGCTCGCGC GCTGTCGGCA 3950
CGTCGGCGCT CAGGCACAGG TTGGACGCCT TGAGGTCCAT GAGCTTGAAC 4000
AGGTAAGCCG TCAGCGGGTG CGTCGCCGTC TCGCTCCTGG CCGCGAAGGT 4050
GGCCTTGAGC GTCGGGTGTG GTGCCATGGC TGATGAGGCT GAGAGAGGCT 4100
GAGGCTGCGG CTGGTTGGAT AGTTTAACCC TTAGGGTGCC GTTGTGGCGG 4150
TTTAGAGGGG GGGAAAAAAA AGAGAGAGAT GGCACAATTC TGCTGTGCGA 4200
ATGACGTTGG AAGCGCGACA GCCGTGCGGG AGGAAGAGGA GTAGGAACTG 4250
TCGGCGATTG GGAGAATTTC GTGCGATCCG AGTCGTCTCG AGGCGAGGGA 4300
GTTGCTTTAA TGTCGGGCTC GTCCCCTGGT CAAAATTCTA GGGAGCAGCG 4350
CTGGCAACGA GAGCAGAGCA GCAGTAGTCG ATGCTAGAAA TCGATAGATC 4400
CACGATGCCA AAAAGCTTGT TCATTTCGGC TAGCCCGTGA TCCTGGCGCT 4450
TCTAGGGCTG AAACTGTGTT GTTAATGTAT TATTGGCTGT GTAACTGACT 4500
TGAATGGGGA ATGAGGAGCG CGATGGATTC GCTTGCATGT CCCCTGGCCA 4550
AGACGAGCCG CTTTGGCGGT TTGTGATTCG AAGGTGTGTC AGCGGAGGCG 4600
CCAGGGCAAC ACGCACTGAG CCAGCCAACA TGCATTGCTG CCGACATGAA 4650
TAGACACGCG CCGAGCAGAC ATAGGAGACG TGTTGACTGT AAAAATTCTA 4700
CTGAATATTA GCACGCATGG TCTCAATAAG AGCAATAGGA ATGCTTGCCA 4750
ATCATAAGTA CGTATGTGCT TTTTCCTGCA AATGGTACGT ACGGACAGTT 4800
CATGTTGTCT GTCATCCCCC ACTCAGGCTC TCATGATCAT TTTATGGGAC 4850
TGGGGTTTTG CTGACTGAAT GGATTCAGCC GCACGAAACA AATTGGGGGC 4900

FIGURE 6C: pTEX2

CATGCAGAAG GGAAGCCCCC CCAGCCCCCT GTTCATAATT TGTTAAGAGT 4950
CGGAGAGCTG CCTAGTATGA AGCAGCAATT GATAACGTTG ACTTTGCGCA 5000
TGAGCTCTGA AGCCGGGCAT ATGTATCACG TTTCTGCCTA GAGCCGCACG 5050
GGACCCAAGA AGCTCTTGTC ATAAGGTATT TATGAGTGTT CAGCTGCCAA 5100
CGCTGGTTCT ACTTTGGCTC AACCGCATCC CATAAGCTGA ACTTTGGGAG 5150
CTGCCAGAAT GTCTCTTGAT GTACAGCGAT CAACAACCGT GCGCCGGTCG 5200
ACAACTGTTC ACCGATCAGG GACGCGAAGA GGACCCAATC CCGGTTAACG 5250
CACCTGCTCC GAAGAAGCAA AAGGGCTATG AGGTGGTGCA GCAAGGAATC 5300
AAAGAGCTCT ATCCACTTGA CAAGGCCAAT GTCGCTCCCG ATCTGGAGTA 5350
AGTCAACCCT GAAGTGGAAG TTTGCTTCTC TGATTAGTAT GTAGCATCGT 5400
GTTTGTCCCA GGACTGGGTG CAAATCCCGA AGACAGCTGG AAGTCCAGCA 5450
AGACCGACTT CAATTGGACC ACGCATACAG ATGGCCTCCA GAGAGACTTC 5500
CCAAGAGCTC GGTTGCTTCT GTATATGTAC GACTCAGCAT GGACTGGCCA 5550
GCTCAAAGTA AAACAATTCA TGGGCAATAT CGCGATGGGG CTCTTGGTTG 5600
GGCTGAGGAG CAAGAGAGAG GTAGGCCAAA CGCCAGACTC GAACCGCCAG 5650
CCAAGTCTCA AACTGACTGC AGGCGGCCGC CATATGCATC CTAGGCCTAT 5700
TAATATTCCG GAGTATACGT AGCCGGCTAA CGTTAACAAC CGGTACCTCT 5750
AGAACTATAG CTAGCATGCG CAAATTTAAA GCGCTGATAT CGATCGCGCG 5800
CAGATCCATA TATAGGGCCC GGGTTATAAT TACCTCAGGT CGACGTCCCA 5850
TGGCCATTCG AATTCGTAAT CATGGTCATA GCTGTTTCCT GTGTGAAATT 5900
GTTATCCGCT CACAATTCCA CACAACATAC GAGCCGGAAG CATAAAGTGT 5950
AAAGCCTGGG GTGCCTAATG AGTGAGCTAA CTCACATTAA TTGCGTTGCG 6000
CTCACTGCCC GCTTTCCAGT CGGGAAACCT GTCGTGCCAG CTGCATTAAT 6050
GAATCGGCCA ACGCGCGGGG AGAGGCGGTT TGCGTATTGG GCGCTCTTCC 6100
GCTTCCTCGC TCACTGACTC GCTGCGCTCG GTCGTTCGGC TGCGGCGAGC 6150
GGTATCAGCT CACTCAAAGG CGGTAATACG GTTATCCACA GAATCAGGGG 6200
ATAACGCAGG AAAGAACATG TGAGCAAAAG GCCAGCAAAA GGCCAGGAAC 6250
CGTAAAAAGG CCGCGTTGCT GGCGTTTTTC CATAGGCTCC GCCCCCCTGA 6300
CGAGCATCAC AAAAATCGAC GCTCAAGTCA GAGGTGGCGA AACCCGACAG 6350
GACTATAAAG ATACCAGGCG TTTCCCCCTG GAAGCTCCCT CGTGCGCTCT 6400
CCTGTTCCGA CCCTGCCGCT TACCGGATAC CTGTCCGCCT TTCTCCCTTC 6450
GGGAAGCGTG GCGCTTTCTC ATAGCTCACG CTGTAGGTAT CTCAGTTCGG 6500
TGTAGGTCGT TCGCTCCAAG CTGGGCTGTG TGCACGAACC CCCCGTTCAG 6550
CCCGACCGCT GCGCCTTATC CGGTAACTAT CGTCTTGAGT CCAACCCGGT 6600
AAGACACGAC TTATCGCCAC TGGCAGCAGC CACTGGTAAC AGGATTAGCA 6650
GAGCGAGGTA TGTAGGCGGT GCTACAGAGT TCTTGAAGTG GTGGCCTAAC 6700
TACGGCTACA CTAGAAGAAC AGTATTTGGT ATCTGCGCTC TGCTGAAGCC 6750
AGTTACCTTC GGAAAAAGAG TTGGTAGCTC TTGATCCGGC AAACAAACCA 6800
CCGCTGGTAG CGGTGGTTTT TTTGTTTGCA AGCAGCAGAT TACGCGCAGA 6850
AAAAAAGGAT CTCAAGAAGA TCCTTTGATC TTTTCTACGG GGTCTGACGC 6900
TCAGTGGAAC GAAAACTCAC GTTAAGGGAT TTTGGTCATG AGATTATCAA 6950
AAAGGATCTT CACCTAGATC CTTTTAAATT AAAAATGAAG TTTTAAATCA 7000
ATCTAAAGTA TATATGAGTA AACTTGGTCT GACAGTTACC AATGCTTAAT 7050
CAGTGAGGCA CCTATCTCAG CGATCTGTCT ATTTCGTTCA TCCATAGTTG 7100
CCTGACTCCC CGTCGTGTAG ATAACTACGA TACGGGAGGG CTTACCATCT 7150
GGCCCCAGTG CTGCAATGAT ACCGCGAGAC CCACGCTCAC CGGCTCCAGA 7200
TTTATCAGCA ATAAACCAGC CAGCCGGAAG GGCCGAGCGC AGAAGTGGTC 7250
CTGCAACTTT ATCCGCCTCC ATCCAGTCTA TTAATTGTTG CCGGGAAGCT 7300
AGAGTAAGTA GTTCGCCAGT TAATAGTTTG CGCAACGTTG TTGCCATTGC 7350

FIGURE 6D: pTEX2

```
TACAGGCATC GTGGTGTCAC GCTCGTCGTT TGGTATGGCT TCATTCAGCT      7400
CCGGTTCCCA ACGATCAAGG CGAGTTACAT GATCCCCCAT GTTGTGCAAA      7450
AAAGCGGTTA GCTCCTTCGG TCCTCCGATC GTTGTCAGAA GTAAGTTGGC      7500
CGCAGTGTTA TCACTCATGG TTATGGCAGC ACTGCATAAT TCTCTTACTG      7550
TCATGCCATC CGTAAGATGC TTTTCTGTGA CTGGTGAGTA CTCAACCAAG      7600
TCATTCTGAG AATAGTGTAT GCGGCGACCG AGTTGCTCTT GCCCGGCGTC      7650
AATACGGGAT AATACCGCGC CACATAGCAG AACTTTAAAA GTGCTCATCA      7700
TTGGAAAACG TTCTTCGGGG CGAAAACTCT CAAGGATCTT ACCGCTGTTG      7750
AGATCCAGTT CGATGTAACC CACTCGTGCA CCCAACTGAT CTTCAGCATC      7800
TTTTACTTTC ACCAGCGTTT CTGGGTGAGC AAAAACAGGA AGGCAAAATG      7850
CCGCAAAAAA GGGAATAAGG GCGACACGGA AATGTTGAAT ACTCATACTC      7900
TTCCTTTTTC AATATTATTG AAGCATTTAT CAGGGTTATT GTCTCATGAG      7950
CGGATACATA TTTGAATGTA TTTAGAAAAA TAAACAAATA GGGGTTCCGC      8000
GCACATTTCC CCGAAAAGTG CCACCTGACG TCTAAGAAAC CATTATTATC      8050
ATGACATTAA CCTATAAAAA TAGGCGTATC ACGAGGCCCT TTCGTCTCGC      8100
GCGTTTCGGT GATGACGGTG AAAACCTCTG ACACATGCAG CTCCCGGAGA      8150
CGGTCACAGC TTGTCTGTAA GCGGATGCCG GGAGCAGACA AGCCCGTCAG      8200
GGCGCGTCAG CGGGTGTTGG CGGGTGTCGG GGCTGGCTTA ACTATGCGGC      8250
ATCAGAGCAG ATTGTACTGA GAGTGCACCA TAAAATTGTA AACGTTAATA      8300
TTTTGTTAAA ATTCGCGTTA AATTTTTGTT AAATCAGCTC ATTTTTTAAC      8350
CAATAGGCCG AAATCGGCAA AATCCCTTAT AAATCAAAAG AATAGCCCGA      8400
GATAGGGTTG AGTGTTGTTC CAGTTTGGAA CAAGAGTCCA CTATTAAAGA      8450
ACGTGGACTC CAACGTCAAA GGGCGAAAAA CCGTCTATCA GGGCGATGGC      8500
CCACTACGTG AACCATCACC CAAATCAAGT TTTTTGGGGT CGAGGTGCCG      8550
TAAAGCACTA AATCGGAACC CTAAAGGGAG CCCCCGATTT AGAGCTTGAC      8600
GGGGAAAGCC GGCGAACGTG GCGAGAAAGG AAGGGAAGAA AGCGAAAGGA      8650
GCGGGCGCTA GGGCGCTGGC AAGTGTAGCG GTCACGCTGC GCGTAACCAC      8700
CACACCCGCC GCGCTTAATG CGCCGCTACA GGGCGCGTAC TATGGTTGCT      8750
TTGACGTATG CGGTGTGAAA TACCGCACAG ATGCGTAAGG AGAAAATACC      8800
GCATCAGGCG CCATTCGCCA TTCAGGCTGC GCAACTGTTG GGAAGGGCGA      8850
TCGGTGCGGG CCTCTTCGCT ATTACGCCAG CTGGCGAAAG GGGGATGTGC      8900
TGCAAGGCGA TTAAGTTGGG TAACGCCAGG GTTTTCCCAG TCACGACGTT      8950
GTAAAACGAC GGCCAGTGCC                                       8970
```

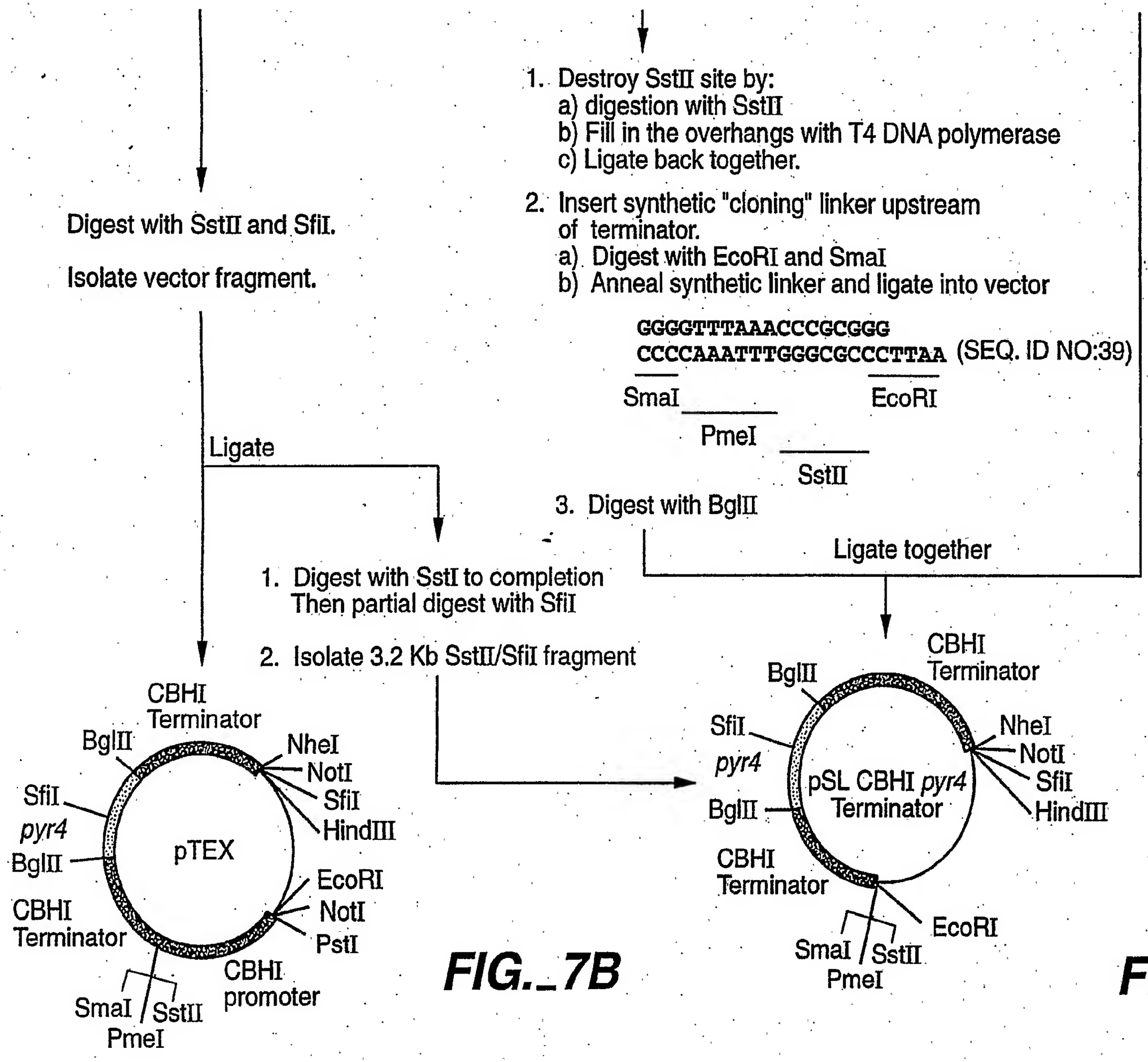


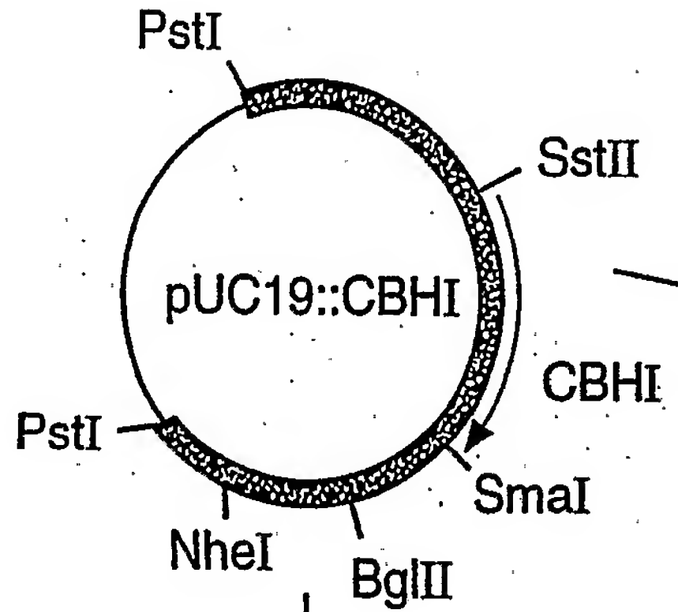
FIG._7B

FIG._7

FIG._7A

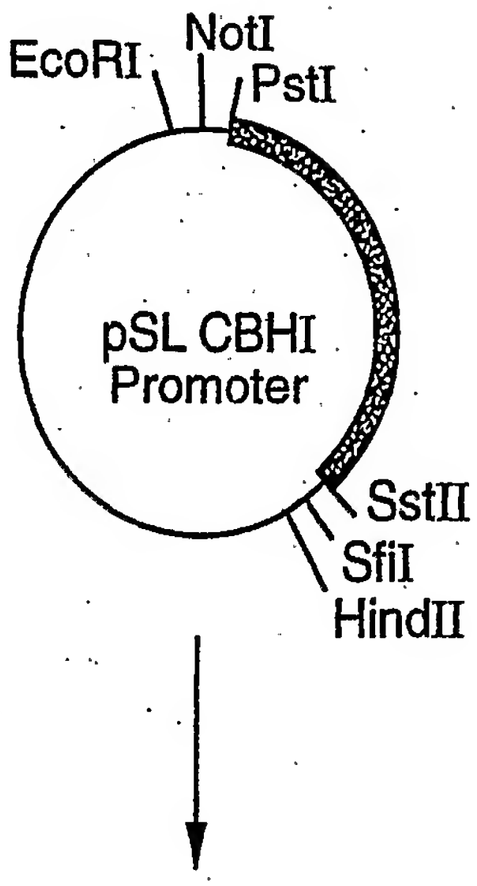
FIG._7B

FIG._7A



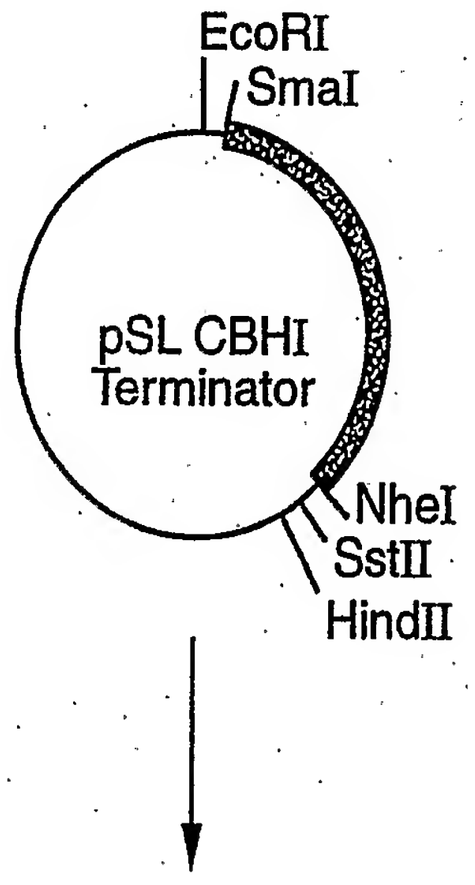
Isolate 5' PstI/SstII
2.2 Kb fragment.

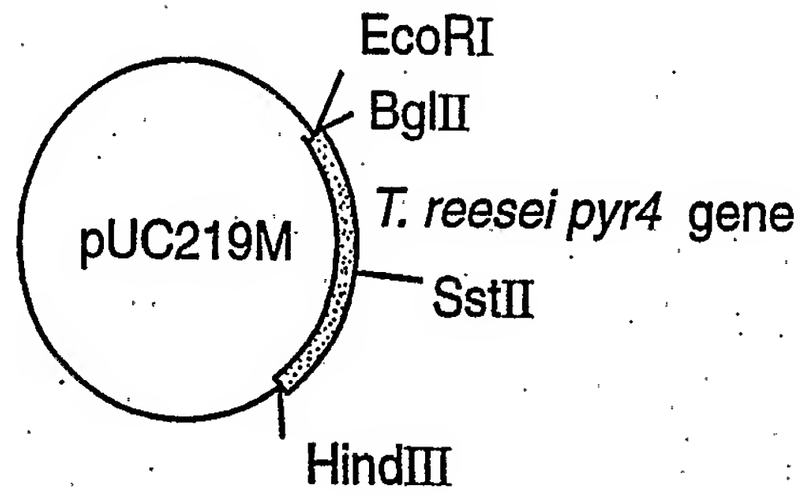
Ligate into PstI/SstII
digested pSL1180 vector



Isolate 3' SmaI/NheI
1.5 Kb fragment.

Ligate into SmaI/NheI
digested pSL1180 vector





1. Destroy the SstII site using site directed mutagenesis

TGAGCCGAGGCCTCC
Sst Kill (SEQ. ID NO:40)

2. Introduce BglII site at HindIII site. Digest with HindIII and ligate in annealed HindIII/BglII/HindIII linkers

AGCTTGAGATCTGA
ACTCTAGACTTCGA (SEQ. ID NO:41)
HindIII BglII HindIII

3. Isolate *pyr4* gene on 1.6 Kb BglII fragment

Figure 8A

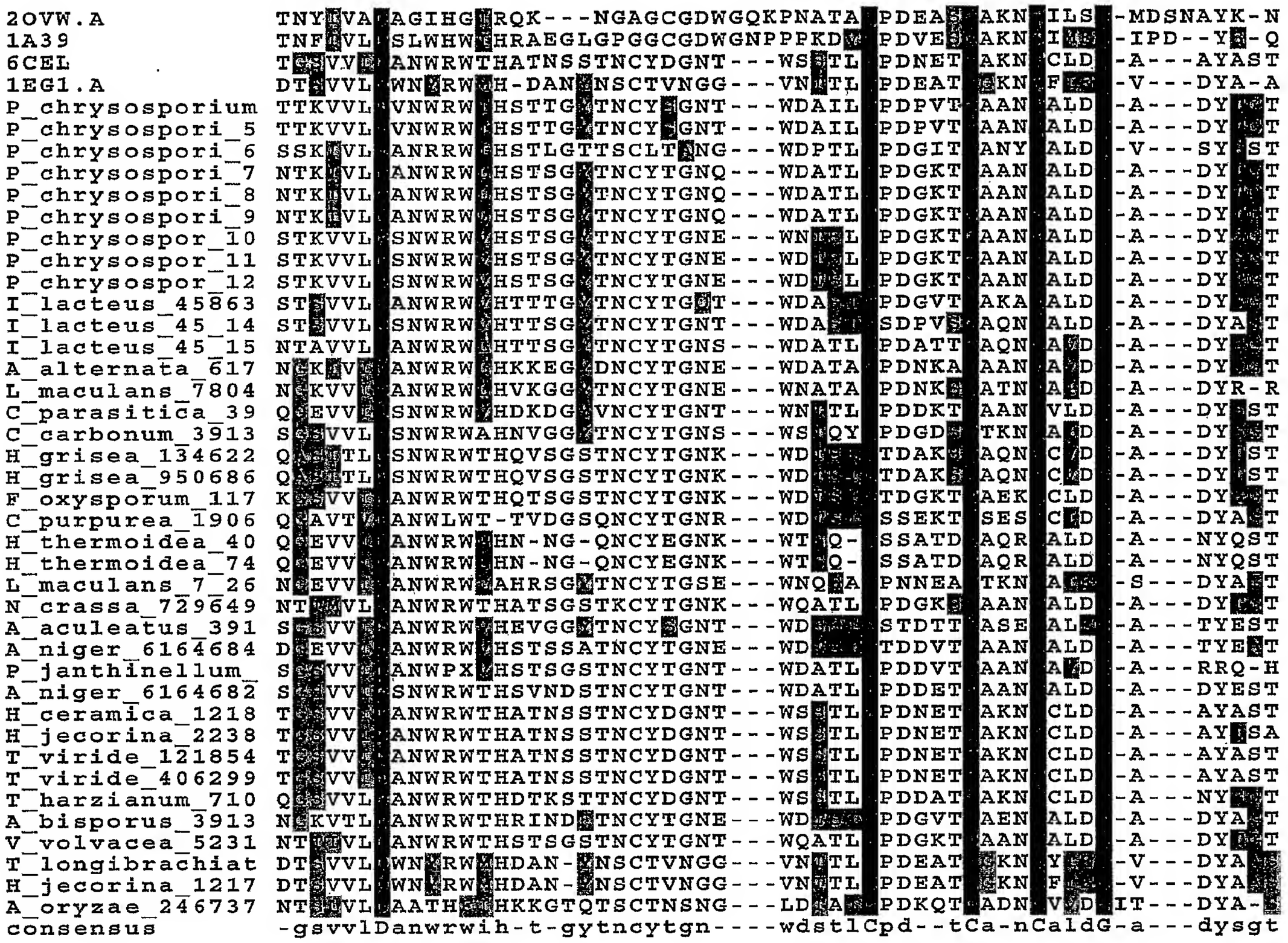


Figure 8B

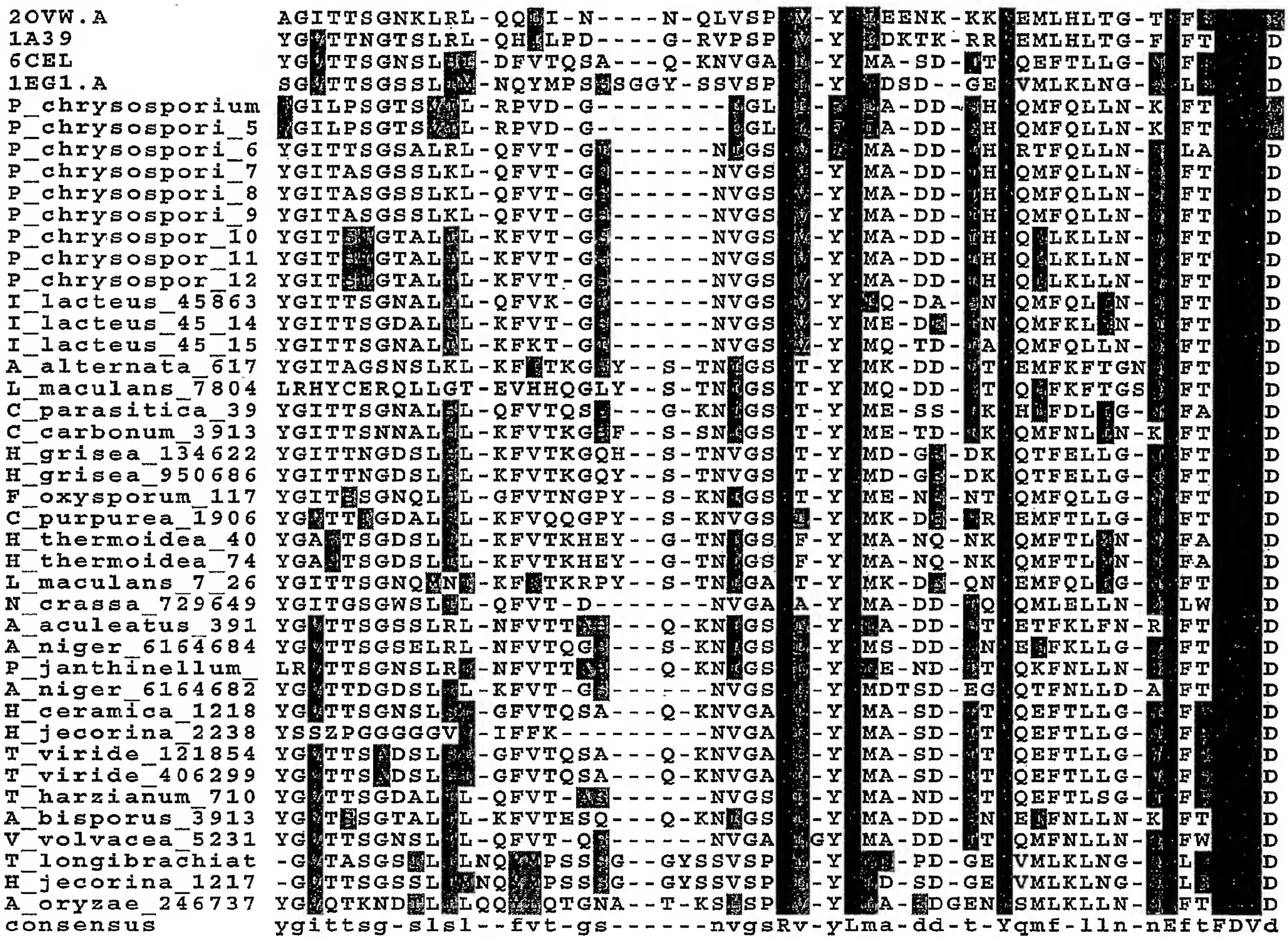

2OVW.A AGITTSGNKLRL-QQ I-N----N-QLVSP -Y EENK-KK EMLHLTG-T F D
1A39 YG TTNGTSLRL-QH LPD----G-RVPSP -Y DKTK-RR EMLHLTG-F FT D
6CEL YG TTSGNSL -DFVTQSA---Q-KNVGA -Y MA-SD- T QEFTLLG- F D
1EG1.A SG TTSGSSL -NQYMPS SGGY-SSVSP -Y DSD--GE VMLKLNG- L D
P_chrysosporium GILPSGTS L-RPVD-G--------GL -A-DD- H QMFQLLN-K FT D
P_chrysospori_5 GILPSGTS L-RPVD-G--------GL -A-DD- H QMFQLLN-K FT D
P_chrysospori_6 YGITTSGSALRL-QFVT-G------NGS -MA-DD- H RTFQLLN- LA D
P_chrysospori_7 YGITASGSSLKL-QFVT-G------NVGS -Y MA-DD- H QMFQLLN- FT D
P_chrysospori_8 YGITASGSSLKL-QFVT-G------NVGS -Y MA-DD- H QMFQLLN- FT D
P_chrysospori_9 YGITASGSSLKL-QFVT-G------NVGS -Y MA-DD- H QMFQLLN- FT D
P_chrysospor_10 YGIT GTAL L-KFVT-G------NVGS -Y MA-DD- H Q LKLLN- FT D
P_chrysospor_11 YGIT GTAL L-KFVT-G------NVGS -Y MA-DD- H Q LKLLN- FT D
P_chrysospor_12 YGIT GTAL L-KFVT-G------NVGS -Y MA-DD- H Q LKLLN- FT D
I_lacteus_45863 YGITTSGNAL L-QFVK-G------NVGS -Y Q-DA- N QMFQL N- FT D
I_lacteus_45_14 YGITTSGDAL L-KFVT-G------NVGS -Y ME-D - N QMFKL N- FT D
I_lacteus_45_15 YGITTSGNAL L-KFKT-G------NVGS -Y MQ-TD- A QMFQLLN- FT D
A_alternata_617 YGITAGSNSLKL-KF TKG Y--S-TN GS T-Y MK-DD- T EMFKFTGN FT D
L_maculans_7804 LRHYCERQLLGT-EVHHQGLY--S-TN GS T-Y MQ-DD- T Q FKFTGS FT D
C_parasitica_39 YGITTSGNAL L-QFVTQS ---G-KN GS T-Y ME-SS- K H FDL G- FA D
C_carbonum_3913 YGITTSNNAL L-KFVTKG F--S-SN GS T-Y ME-TD- K QMFNL N-K FT D
H_grisea_134622 YGITTNGDSL L-KFVTKGQH--S-TNVGS T-Y MD-G -DK QTFELLG- FT D
H_grisea_950686 YGITTNGDSL L-KFVTKGQY--S-TNVGS T-Y MD-G -DK QTFELLG- FT D
F_oxysporum_117 YGIT SGNQL L-GFVTNGPY--S-KN GS T-Y ME-N -NT QMFQLLG- FT D
C_purpurea_1906 YG TT GDAL L-KFVQQGPY--S-KNVGS -Y MK-D -R EMFTLLG- FT D
H_thermoidea_40 YGA TSGDSL L-KFVTKHEY--G-TN GS F-Y MA-NQ-NK QMFTL N- FA D
H_thermoidea_74 YGA TSGDSL L-KFVTKHEY--G-TN GS F-Y MA-NQ-NK QMFTL N- FA D
L_maculans_7_26 YGITTSGNQ N -KF TKRPY--S-TN GA T-Y MK-D -QN EMFQL G- FT D
N_crassa_729649 YGITGSGWSL L-QFVT-D-------NVGA A-Y MA-DD- Q QMLELLN- LW D
A_aculeatus_391 YG TTSGSSLRL-NFVTT ---Q-KN GS -Y A-DD- T ETFKLFN-R FT D
A_niger_6164684 YG TTSGSELRL-NFVTQG ---S-KN GS L-Y MS-DD- N E FKLLG- FT D
P_janthinellum_ LR TTSGNSLR -NFVTT ---Q-KN GS -Y E-ND- T QKFNLLN- FT D
A_niger_6164682 YG TTDGDSL L-KFVT-G-------NVGS -Y MDTSD-EG QTFNLLD-A FT D
H_ceramica_1218 YG TTSGNSL -GFVTQSA---Q-KNVGA -Y MA-SD- T QEFTLLG- F D
H_jecorina_2238 YSSZPGGGGGV -IFFK---------NVGA -Y MA-SD- T QEFTLLG- F D
T_viride_121854 YG TTS DSL -GFVTQSA---Q-KNVGA -Y MA-SD- T QEFTLLG- F D
T_viride_406299 YG TTS DSL -GFVTQSA---Q-KNVGA -Y MA-SD- T QEFTLLG- F D
T_harzianum_710 YG TTSGDAL L-QFVT- S------NVGS -Y MA-ND- T QEFTLSG- F D
A_bisporus_3913 YG T SGTAL L-KFVTESQ---Q-KN GS -Y MA-DD- N E FNLLN-K FT D
V_volvacea_5231 YG TTSGNSL L-QFVT-Q -------NVGA GY MA-DD- T QMFNLLN- FW D
T_longibrachiat -G TASGS L LNQ PSS G--GYSSVSP -Y -PD-GE VMLKLNG- L D
H_jecorina_1217 -G TTSGSSL NQ PSS G--GYSSVSP -Y D-SD-GE VMLKLNG- L D
A_oryzae_246737 YG QTKND L LQQ QTGNA--T-KS SP -Y A- DGEN SMLKLLN- FT D
consensus ygittsg-slsl--fvt-gs------nvgsRv-yLma-dd-t-Yqmf-lln-nEftFDVd

Figure 8C

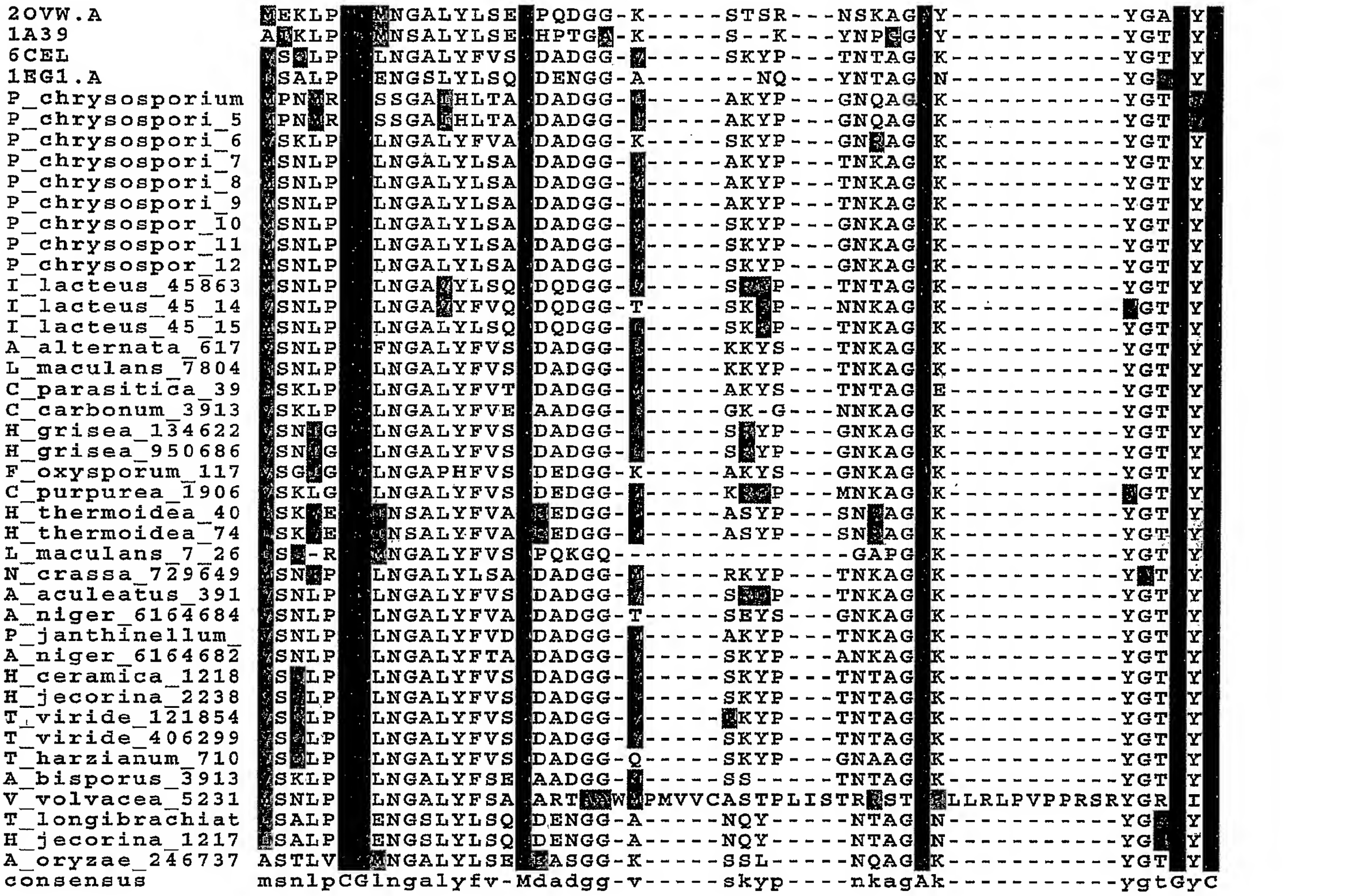

2OVW.A
1A39
6CEL
1EG1.A
P_chrysosporium
P_chrysospori_5
P_chrysospori_6
P_chrysospori_7
P_chrysospori_8
P_chrysospori_9
P_chrysospor_10
P_chrysospor_11
P_chrysospor_12
I_lacteus_45863
I_lacteus_45_14
I_lacteus_45_15
A_alternata_617
L_maculans_7804
C_parasitica_39
C_carbonum_3913
H_grisea_134622
H_grisea_950686
F_oxysporum_117
C_purpurea_1906
H_thermoidea_40
H_thermoidea_74
L_maculans_7_26
N_crassa_729649
A_aculeatus_391
A_niger_6164684
P_janthinellum_
A_niger_6164682
H_ceramica_1218
H_jecorina_2238
T_viride_121854
T_viride_406299
T_harzianum_710
A_bisporus_3913
V_volvacea_5231
T_longibrachiat
H_jecorina_1217
A_oryzae_246737
consensus

Figure 8D

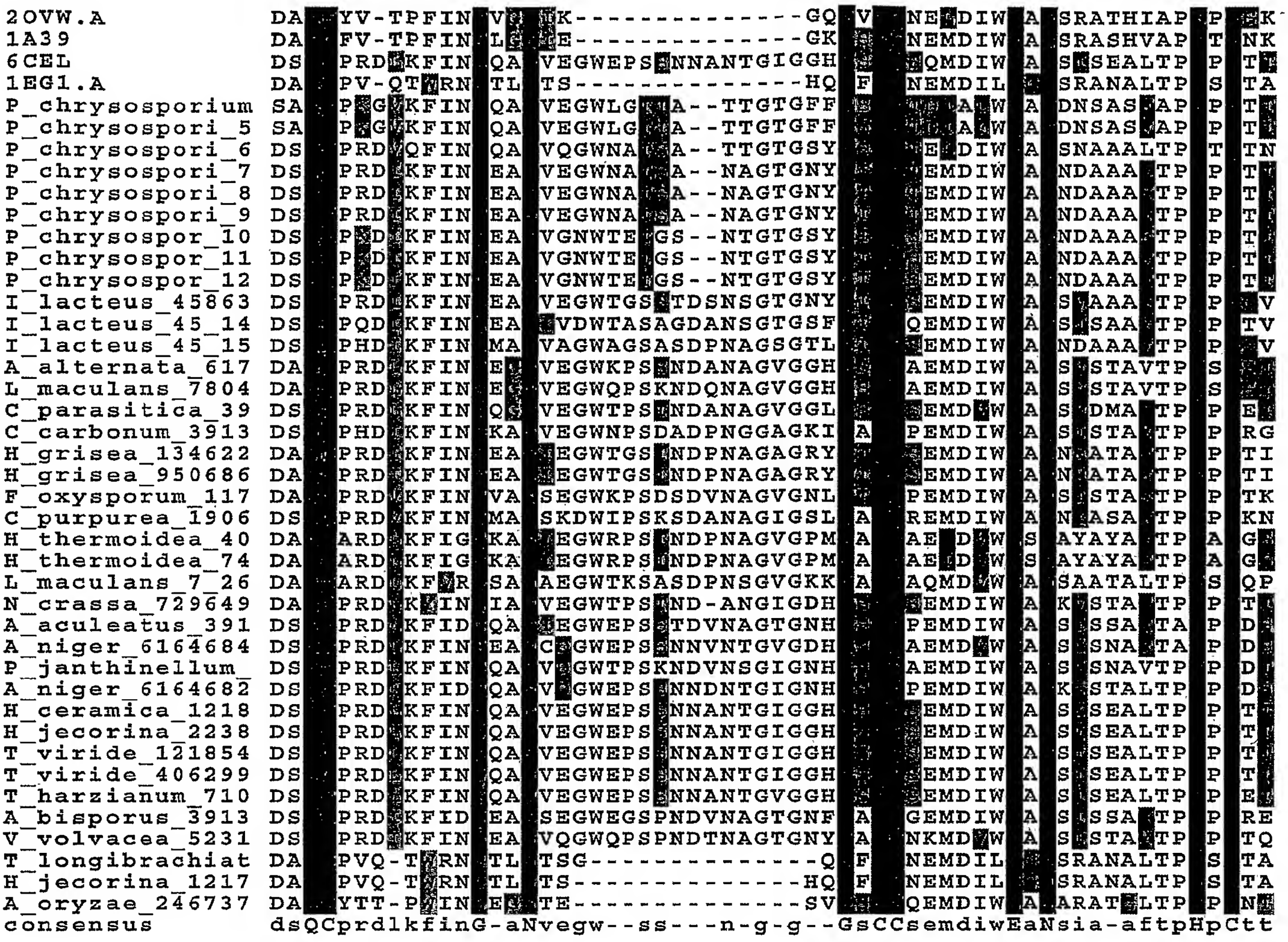

2OVW.A DA YV-TPFIN VQ K----------GQ V NE DIW A SRATHIAP P K
1A39 DA FV-TPFIN LG E----------GK NEMDIW A SRASHVAP T NK
6CEL DS PRD KFIN QA VEGWEPS NNANTGIGGH QMDIW A S SEALTP P T
1EG1.A DA PV-QT RN TL TS----------HQ F NEMDIL SRANALTP S TA
P_chrysosporium SA P G KFIN QA VEGWLG A--TTGTGFF AW A DNSAS AP P T
P_chrysospori_5 SA P G KFIN QA VEGWLG A--TTGTGFF AW A DNSAS AP P T
P_chrysospori_6 DS PRD QFIN QA VQGWNA A--TTGTGSY E DIW A SNAAALTP T TN
P_chrysospori_7 DS PRD KFIN EA VEGWNA A--NAGTGNY EMDIW A NDAAA TP P T
P_chrysospori_8 DS PRD KFIN EA VEGWNA A--NAGTGNY EMDIW A NDAAA TP P T
P_chrysospori_9 DS PRD KFIN EA VEGWNA A--NAGTGNY EMDIW A NDAAA TP P T
P_chrysospor_10 DS P D KFIN EA VGNWTE GS--NTGTGSY EMDIW A NDAAA TP P T
P_chrysospor_11 DS P D KFIN EA VGNWTE GS--NTGTGSY EMDIW A NDAAA TP P T
P_chrysospor_12 DS P D KFIN EA VGNWTE GS--NTGTGSY EMDIW A NDAAA TP P T
I_lacteus_45863 DS PRD KFIN EA VEGWTGS TDSNSGTGNY EMDIW A S AAA TP P V
I_lacteus_45_14 DS PQD KFIN EA VDWTASAGDANSGTGSF QEMDIW A S SAA TP P TV
I_lacteus_45_15 DS PHD KFIN MA VAGWAGSASDPNAGSGTL EMDIW A NDAAA TP P V
A_alternata_617 DA PRD KFIN E VEGWKPS NDANAGVGGH AEMDIW A S STAVTP S
L_maculans_7804 DA PRD KFIN E VEGWQPSKNDQNAGVGGH AEMDIW A S STAVTP S
C_parasitica_39 DS PRD KFIN Q VEGWTPS NDANAGVGGL EMD W A S DMA TP P E
C_carbonum_3913 DS PHD KFIN KA VEGWNPSDADPNGGAGKI A PEMDIW A S STA TP P RG
H_grisea_134622 DA PRD KFIN EA EGWTGS NDPNAGAGRY EMDIW A N ATA TP P TI
H_grisea_950686 DA PRD KFIN EA EGWTGS NDPNAGAGRY EMDIW A N ATA TP P TI
F_oxysporum_117 DA PRD KFIN VA SEGWKPSDSDVNAGVGNL PEMDIW A S STA TP P TK
C_purpurea_1906 DS PRD KFIN MA SKDWIPSKSDANAGIGSL A REMDIW A N ASA TP P KN
H_thermoidea_40 DA ARD KFIG KA EGWRPS NDPNAGVGPM A AE D W S AYAYA TP A G
H_thermoidea_74 DA ARD KFIG KA EGWRPS NDPNAGVGPM A AE D W S AYAYA TP A G
L_maculans_7_26 DA ARD KF R SA AEGWTKSASDPNSGVGKK A AQMD W A SAATALTP S QP
N_crassa_729649 DA PRD K IN IA VEGWTPS ND-ANGIGDH EMDIW A K STA TP P T
A_aculeatus_391 DS PRD KFID QA EGWEPS TDVNAGTGNH PEMDIW A S SSA TA P D
A_niger_6164684 DS PRD KFIN EA C GWEPS NNVNTGVGDH AEMD W A S SNA TA P D
P_janthinellum_ DS PRD KFIN QA V GWTPSKNDVNSGIGNH AEMDIW A S SNAVTP P D
A_niger_6164682 DS PRD KFID QA V GWEPS NNDNTGIGNH PEMDIW A K STALTP P D
H_ceramica_1218 DS PRD KFIN QA VEGWEPS NNANTGIGGH EMDIW A S SEALTP P T
H_jecorina_2238 DS PRD KFIN QA VEGWEPS NNANTGIGGH EMDIW A S SEALTP P T
T_viride_121854 DS PRD KFIN QA VEGWEPS NNANTGIGGH EMDIW A S SEALTP P T
T_viride_406299 DS PRD KFIN QA VEGWEPS NNANTGIGGH EMDIW A S SEALTP P T
T_harzianum_710 DS PRD KFIN QA VEGWEPS NNANTGVGGH EMDIW A S SEALTP P E
A_bisporus_3913 DS PRD KFID EA SEGWEGSPNDVNAGTGNF A GEMDIW A S SSA TP P RE
V_volvacea_5231 DS PRD KFIN EA VQGWQPSPNDTNAGTGNY A NKMD W A S STA TP P TQ
T_longibrachiat DA PVQ-T RN TL TSG----------Q F NEMDIL SRANALTP S TA
H_jecorina_1217 DA PVQ-T RN TL TS----------HQ F NEMDIL SRANALTP S TA
A_oryzae_246737 DA YTT-P IN E TE----------SV QEMDIW A ARAT LTP P N
consensus dsQCprdlkfinG-aNvegw--ss---n-g-g--GsCCsemdiwEaNsia-aftpHpCtt

Figure 8E

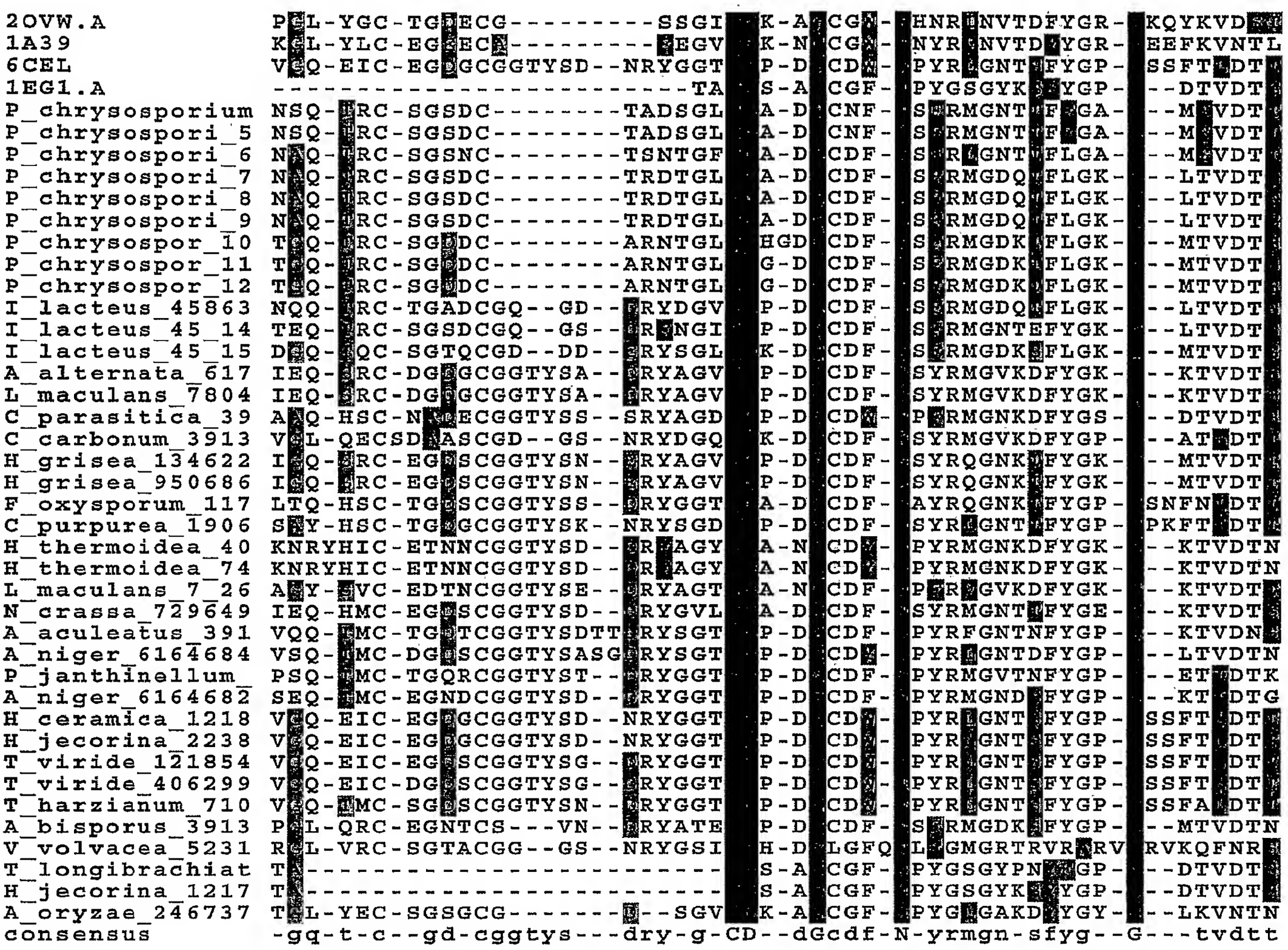

2OVW.A
1A39
6CEL
1EG1.A
P_chrysosporium
P_chrysospori_5
P_chrysospori_6
P_chrysospori_7
P_chrysospori_8
P_chrysospori_9
P_chrysospor_10
P_chrysospor_11
P_chrysospor_12
I_lacteus_45863
I_lacteus_45_14
I_lacteus_45_15
A_alternata_617
L_maculans_7804
C_parasitica_39
C_carbonum_3913
H_grisea_134622
H_grisea_950686
F_oxysporum_117
C_purpurea_1906
H_thermoidea_40
H_thermoidea_74
L_maculans_7_26
N_crassa_729649
A_aculeatus_391
A_niger_6164684
P_janthinellum_
A_niger_6164682
H_ceramica_1218
H_jecorina_2238
T_viride_121854
T_viride_406299
T_harzianum_710
A_bisporus_3913
V_volvacea_5231
T_longibrachiat
H_jecorina_1217
A_oryzae_246737
consensus

Figure 8F

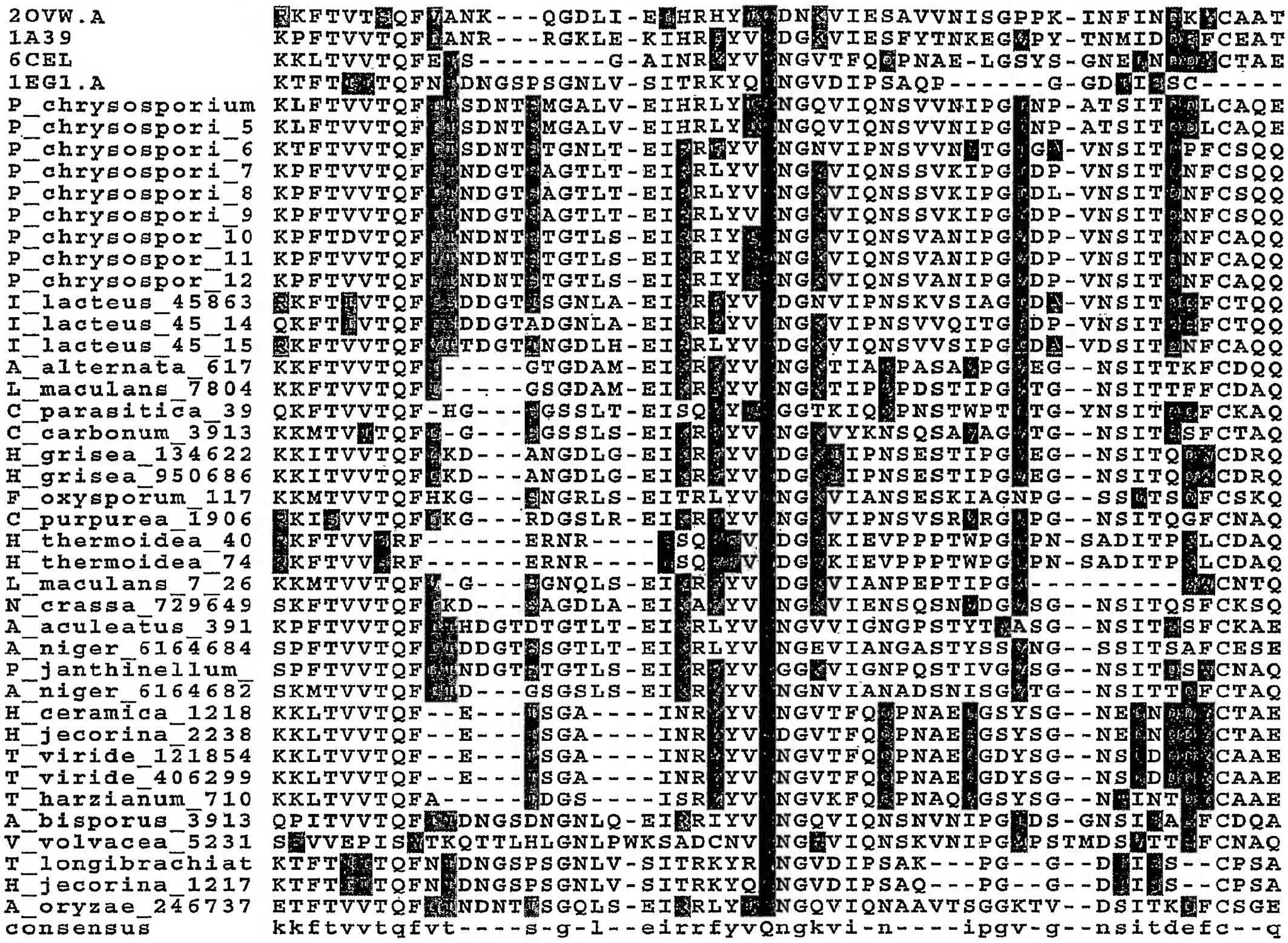


Figure 8G

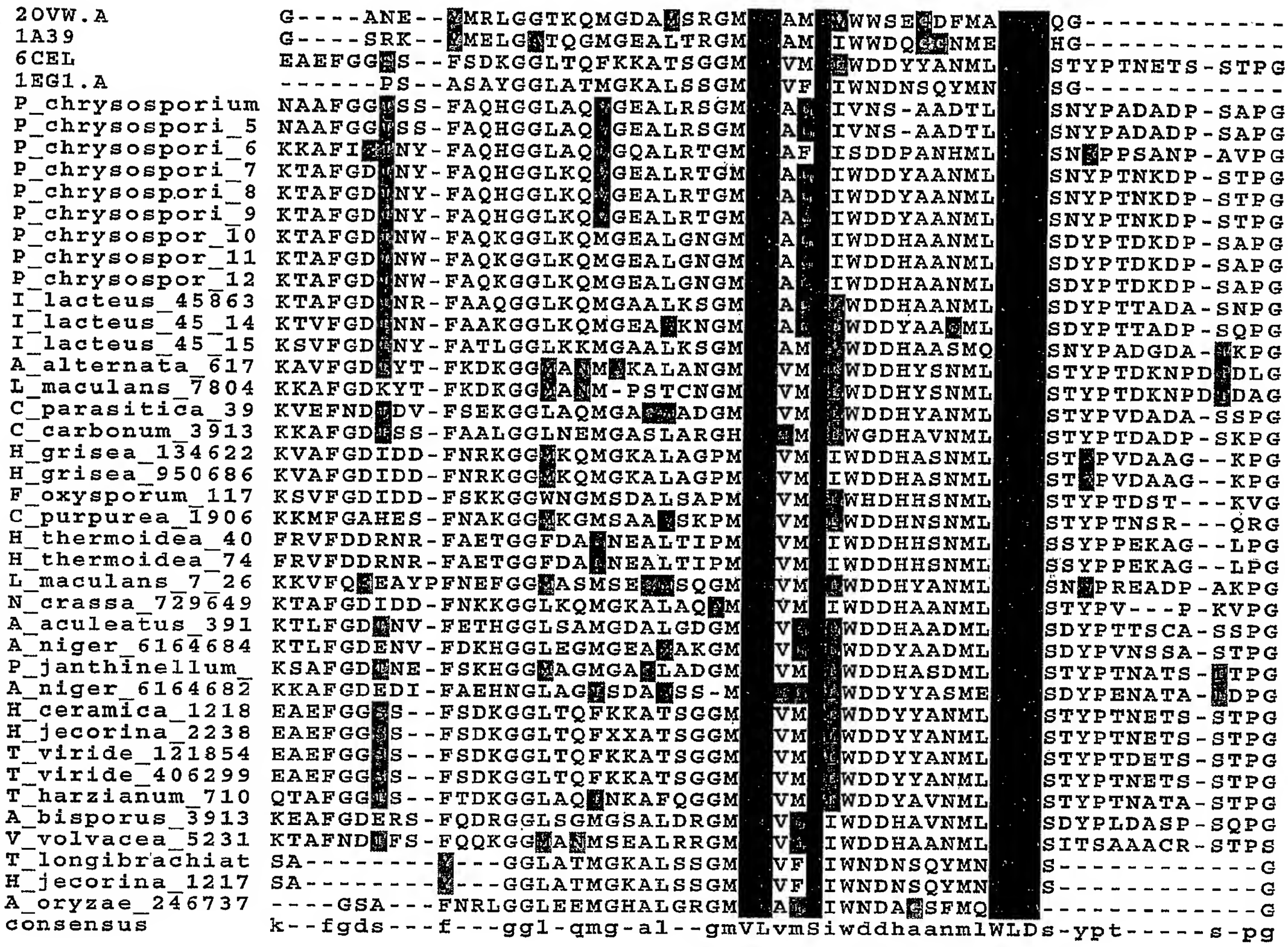


Figure 8H

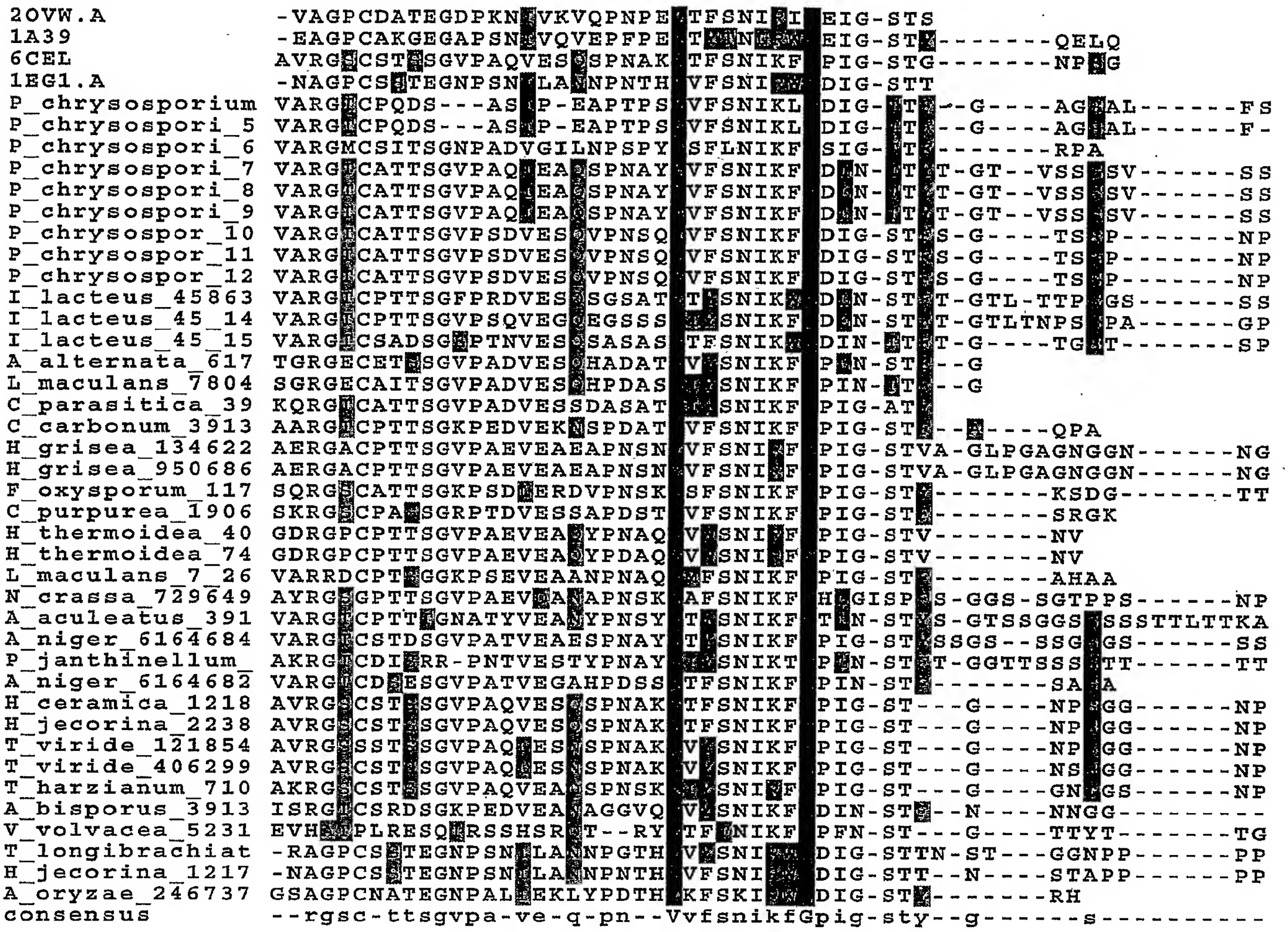

2OVW.A -VAGPCDATEGDPKN VKVQPNPE TFSNI I EIG-STS
1A39 -EAGPCAKGEGAPSN VQVEPFPE T N EIG-ST -------QELQ
6CEL AVRG CST SGVPAQVES SPNAK TFSNIKF PIG-STG -------NP G
1EG1.A -NAGPCS TEGNPSN LA NPNTH VFSNI DIG-STT
P_chrysosporium VARG CPQDS---AS P-EAPTPS VFSNIKL DIG- T --G----AG AL------FS
P_chrysospori_5 VARG CPQDS---AS P-EAPTPS VFSNIKL DIG- T --G----AG AL------F-
P_chrysospori_6 VARGMCSITSGNPADVGILNPSPY SFLNIKF SIG- T -------RPA
P_chrysospori_7 VARG CATTSGVPAQ EA SPNAY VFSNIKF D N- T T-GT--VSS SV------SS
P_chrysospori_8 VARG CATTSGVPAQ EA SPNAY VFSNIKF D N- T T-GT--VSS SV------SS
P_chrysospori_9 VARG CATTSGVPAQ EA SPNAY VFSNIKF D N- T T-GT--VSS SV------SS
P_chrysospor_10 VARG CATTSGVPSDVES VPNSQ VFSNIKF DIG-ST S-G----TS P-------NP
P_chrysospor_11 VARG CATTSGVPSDVES VPNSQ VFSNIKF DIG-ST S-G----TS P-------NP
P_chrysospor_12 VARG CATTSGVPSDVES VPNSQ VFSNIKF DIG-ST S-G----TS P-------NP
I_lacteus_45863 VARG CPTTSGFPRDVES SGSAT T SNIK D N-ST T-GTL-TTP GS------SS
I_lacteus_45_14 VARG CPTTSGVPSQVEG EGSSS SNIKF D N-ST T-GTLTNPS PA------GP
I_lacteus_45_15 VARG CSADSG PTNVES SASAS TFSNIK DIN- T T-G----TG T-------SP
A_alternata_617 TGRGECET SGVPADVES HADAT V SNIKF P N-ST --G
L_maculans_7804 SGRGECAITSGVPADVES HPDAS SNIKF PIN- T --G
C_parasitica_39 KQRG CATTSGVPADVESSDASAT SNIKF PIG-AT
C_carbonum_3913 AARG CPTTSGKPEDVEK SPDAT VFSNIKF PIG-ST --A----QPA
H_grisea_134622 AERGACPTTSGVPAEVEAEAPNSN VFSNI F PIG-STVA-GLPGAGNGGN------NG
H_grisea_950686 AERGACPTTSGVPAEVEAEAPNSN VFSNI F PIG-STVA-GLPGAGNGGN------NG
F_oxysporum_117 SQRG CATTSGKPSD ERDVPNSK SFSNIKF PIG-ST -------KSDG-------TT
C_purpurea_1906 SKRG CPA SGRPTDVESSAPDST VFSNIKF PIG-ST -------SRGK
H_thermoidea_40 GDRGPCPTTSGVPAEVEA YPNAQ V SNI F PIG-STV-------NV
H_thermoidea_74 GDRGPCPTTSGVPAEVEA YPDAQ V SNI F PIG-STV-------NV
L_maculans_7_26 VARRDCPT GGKPSEVEAANPNAQ FSNIKF PIG-ST -------AHAA
N_crassa_729649 AYRG GPTTSGVPAEV A APNSK AFSNIKF H GISP S-GGS-SGTPPS------NP
A_aculeatus_391 VARG CPTT GNATYVEA YPNSY T SNIKF T N-ST S-GTSSGGS SSSTTLTTKA
A_niger_6164684 VARG CSTDSGVPATVEAESPNAY T SNIKF PIG-ST SSGS--SSG GS------SS
P_janthinellum_ AKRG CDI RR-PNTVESTYPNAY SNIKT P N-ST T-GGTTSSS TT------TT
A_niger_6164682 VARG CD ESGVPATVEGAHPDSS TFSNIKF PIN-ST -------SA A
H_ceramica_1218 AVRG CST SGVPAQVES SPNAK TFSNIKF PIG-ST ---G----NP GG------NP
H_jecorina_2238 AVRG CST SGVPAQVES SPNAK TFSNIKF PIG-ST ---G----NP GG------NP
T_viride_121854 AVRG SST SGVPAQ ES SPNAK V SNIKF PIG-ST ---G----NP GG------NP
T_viride_406299 AVRG CST SGVPAQ ES SPNAK V SNIKF PIG-ST ---G----NS GG------NP
T_harzianum_710 AKRG CST SGVPAQVEA SPNSK SNI F PIG-ST ---G----GN GS------NP
A_bisporus_3913 ISRG CSRDSGKPEDVEA AGGVQ V SNIKF DIN-ST --N----NNGG---------
V_volvacea_5231 EVH PLRESQ RSSHSR T--RY TF NIKF PFN-ST ---G----TTYT-------TG
T_longibrachiat -RAGPCS TEGNPSN LA NPGTH V SNI DIG-STTN-ST ---GGNPP------PP
H_jecorina_1217 -NAGPCS TEGNPSN LA NPNTH VFSNI DIG-STT--N----STAPP------PP
A_oryzae_246737 GSAGPCNATEGNPAL EKLYPDTH KFSKI DIG-ST -------RH
consensus --rgsc-ttsgvpa-ve-q-pn--Vvfsnikf Gpig-sty--g------s----------

Figure 8I

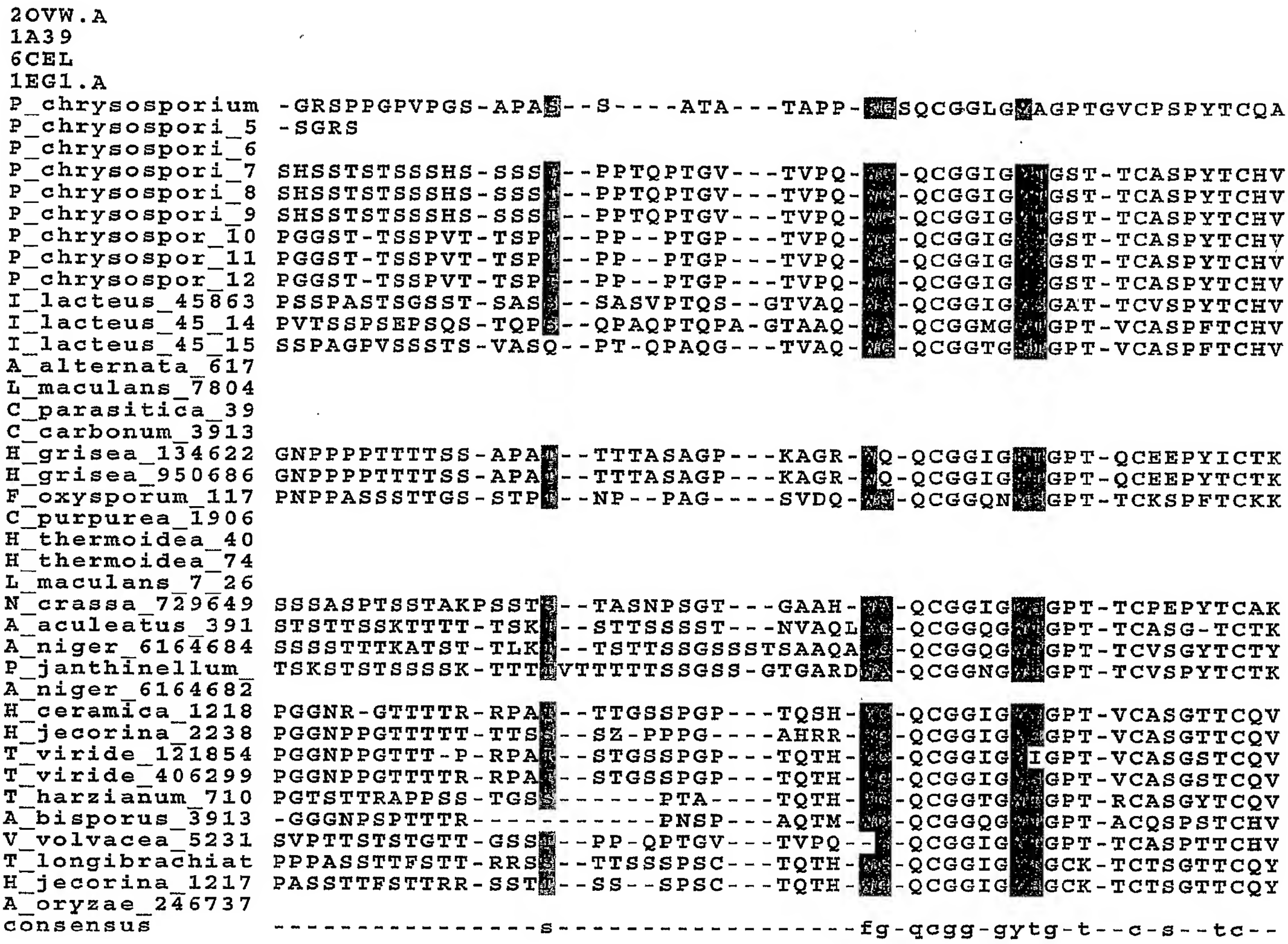
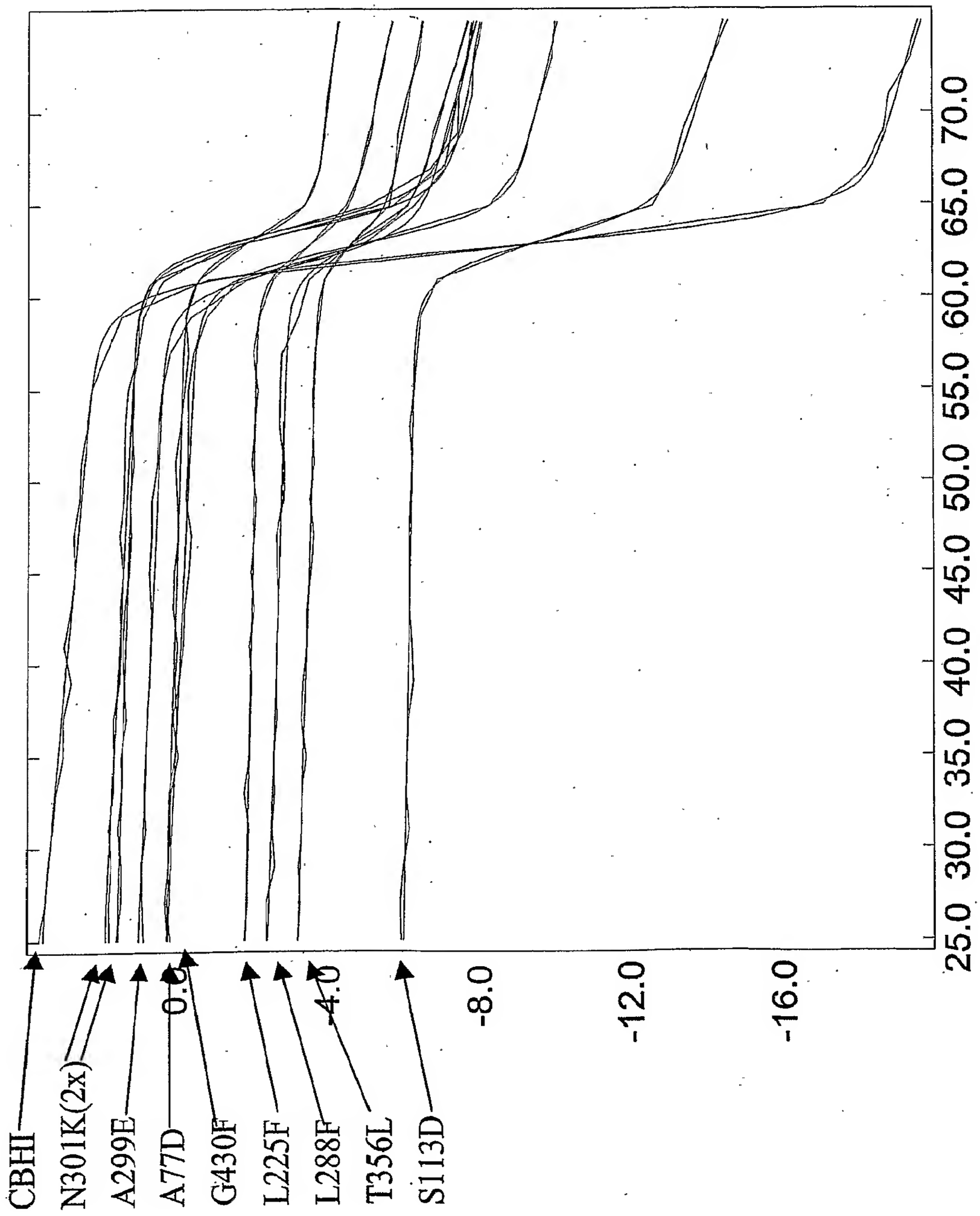

2OVW.A
1A39
6CEL
1EG1.A
P_chrysosporium -GRSPPGPVPGS-APAS--S----ATA---TAPP-SQCGGLGAGPTGVCPSPYTCQA
P_chrysospori_5 -SGRS
P_chrysospori_6
P_chrysospori_7 SHSSTSTSSSHS-SSS--PPTQPTGV---TVPQ--QCGGIGGST-TCASPYTCHV
P_chrysospori_8 SHSSTSTSSSHS-SSS--PPTQPTGV---TVPQ--QCGGIGGST-TCASPYTCHV
P_chrysospori_9 SHSSTSTSSSHS-SSS--PPTQPTGV---TVPQ--QCGGIGGST-TCASPYTCHV
P_chrysospor_10 PGGST-TSSPVT-TSP--PP--PTGP---TVPQ--QCGGIGGST-TCASPYTCHV
P_chrysospor_11 PGGST-TSSPVT-TSP--PP--PTGP---TVPQ--QCGGIGGST-TCASPYTCHV
P_chrysospor_12 PGGST-TSSPVT-TSP--PP--PTGP---TVPQ--QCGGIGGST-TCASPYTCHV
I_lacteus_45863 PSSPASTSGSST-SAS--SASVPTQS--GTVAQ--QCGGIGGAT-TCVSPYTCHV
I_lacteus_45_14 PVTSSPSEPSQS-TQP--QPAQPTQPA-GTAAQ--QCGGMGGPT-VCASPFTCHV
I_lacteus_45_15 SSPAGPVSSSTS-VASQ--PT-QPAQG---TVAQ--QCGGTGGPT-VCASPFTCHV
A_alternata_617
L_maculans_7804
C_parasitica_39
C_carbonum_3913
H_grisea_134622 GNPPPPTTTTSS-APA--TTTASAGP---KAGR-Q-QCGGIGGPT-QCEEPYICTK
H_grisea_950686 GNPPPPTTTTSS-APA--TTTASAGP---KAGR-Q-QCGGIGGPT-QCEEPYTCTK
F_oxysporum_117 PNPPASSSTTGS-STP--NP--PAG----SVDQ--QCGGQNGPT-TCKSPFTCKK
C_purpurea_1906
H_thermoidea_40
H_thermoidea_74
L_maculans_7_26
N_crassa_729649 SSSASPTSSTAKPSST--TASNPSGT---GAAH--QCGGIGGPT-TCPEPYTCAK
A_aculeatus_391 STSTTSSKTTTT-TSK--STTSSSST---NVAQL-QCGGQGGPT-TCASG-TCTK
A_niger_6164684 SSSSTTTKATST-TLK--TSTTSSGSSSTSAAQA-QCGGQGGPT-TCVSGYTCTY
P_janthinellum_ TSKSTSTSSSSK-TTTTVTTTTTSSGSS-GTGARD-QCGGNGGPT-TCVSPYTCTK
A_niger_6164682
H_ceramica_1218 PGGNR-GTTTTR-RPA--TTGSSPGP---TQSH--QCGGIGGPT-VCASGTTCQV
H_jecorina_2238 PGGNPPGTTTTT-TTS--SZ-PPPG----AHRR--QCGGIGGPT-VCASGTTCQV
T_viride_121854 PGGNPPGTTT-P-RPA--STGSSPGP---TQTH--QCGGIGIGPT-VCASGSTCQV
T_viride_406299 PGGNPPGTTTTR-RPA--STGSSPGP---TQTH--QCGGIGGPT-VCASGSTCQV
T_harzianum_710 PGTSTTRAPPSS-TGS-------PTA----TQTH--QCGGTGGPT-RCASGYTCQV
A_bisporus_3913 -GGGNPSPTTTR----------PNSP---AQTM--QCGGQGGPT-ACQSPSTCHV
V_volvacea_5231 SVPTTSTSTGTT-GSS--PP-QPTGV---TVPQ---QCGGIGGPT-TCASPTTCHV
T_longibrachiat PPPASSTTFSTT-RRS--TTSSSPSC---TQTH--QCGGIGGCK-TCTSGTTCQY
H_jecorina_1217 PASSTTFSTTRR-SST--SS--SPSC---TQTH--QCGGIGGCK-TCTSGTTCQY
A_oryzae_246737
consensus ----------------s--------------------fg-qcgg-gytg-t--c-s--tc--

Figure 8J

```
2OVW.A
1A39
6CEL
1EG1.A
P_chrysosporium LNIWYSQ-CI
P_chrysospori_5
P_chrysospori_6
P_chrysospori_7 LNPWYSQ-CY
P_chrysospori_8 LNPWYSQ-CY
P_chrysospori_9 LNPWYSQ-CY
P_chrysospor_10 LNPWYSQ-CY
P_chrysospor_11 LNPWYSQ-CY
P_chrysospor_12 LNPCESILSLQRSSNADQYLQTTRSATKRRLDTALQPRK
I_lacteus_45863 VNAWYSQ-CY
I_lacteus_45_14 LNPWYSQ-CY
I_lacteus_45_15 VNPWYSQ-CY
A_alternata_617
L_maculans_7804
C_parasitica_39
C_carbonum_3913
H_grisea_134622 LNDWYSQ-CL
H_grisea_950686 LNDWYSQ-CL
F_oxysporum_117 INDWYSQ-CQ
C_purpurea_1906
H_thermoidea_40
H_thermoidea_74
L_maculans_7_26
N_crassa_729649 DHDIYSQ-CV
A_aculeatus_391 QNDWYSQ-CL
A_niger_6164684 ENAWYSQ-CL
P_janthinellum_ QNDWYSQ-CL
A_niger_6164682
H_ceramica_1218 LNPWYSQ-CL
H_jecorina_2238 LNPWYSQ-CL
T_viride_121854 LNPWYSQ-CL
T_viride_406299 LNPWYSQ-CL
T_harzianum_710 LNPWYSQ-CL
A_bisporus_3913 INDWYSQ-CF
V_volvacea_5231 LNPWYSQ-CY
T_longibrachiat GNDWYSQ-CL
H_jecorina_1217 SNDWYSQ-CL
A_oryzae_246737
consensus       -n-yysq-c-
```

Figure 8K

Figure 9A


CBHI
N301K(2x)
A299E
A77D
G430F
L225F
L288F
T356L
S113D
0.0
-4.0
-8.0
-12.0
-16.0
25.0
30.0
35.0
40.0
45.0
50.0
55.0
60.0
65.0
70.0

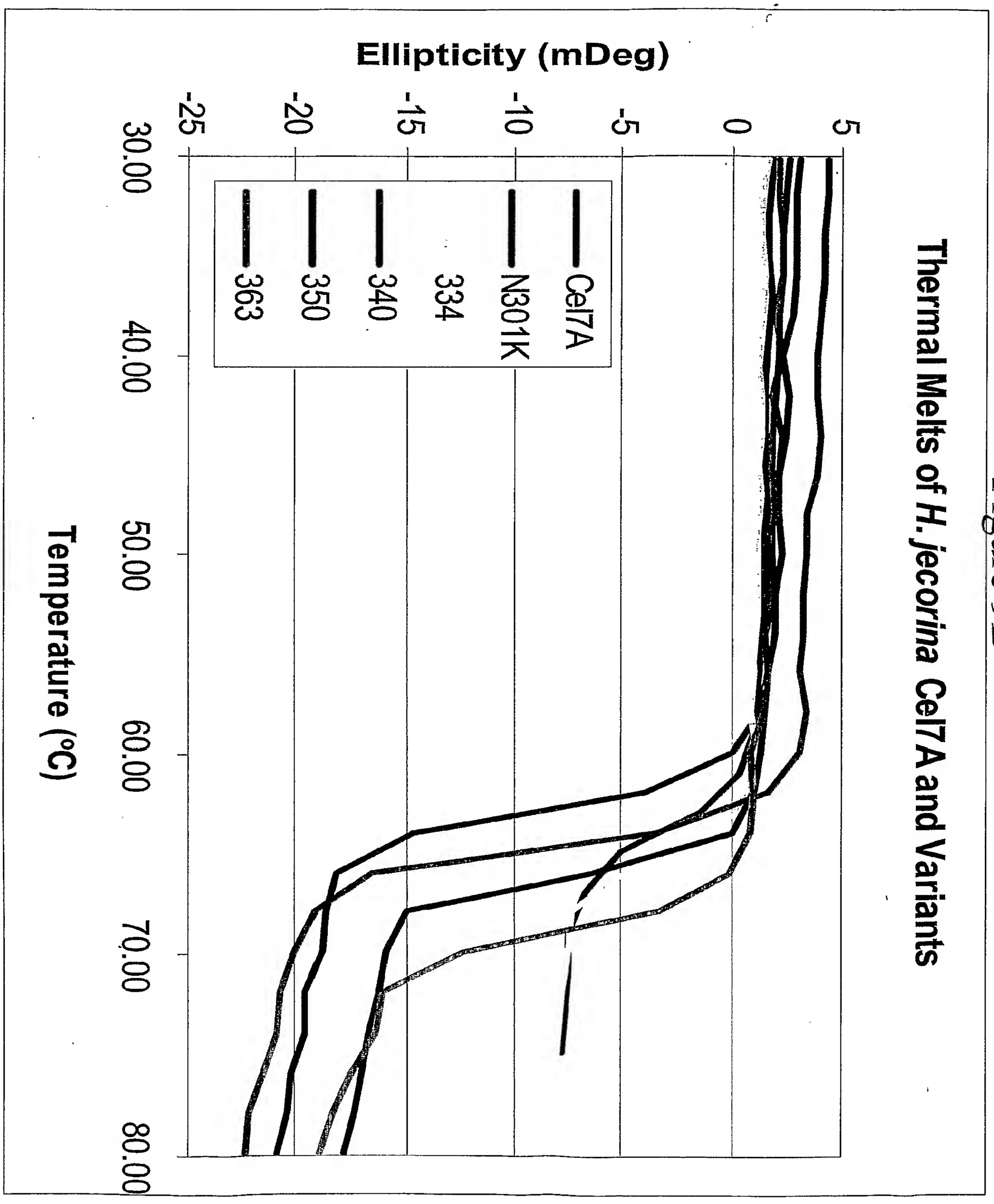

Ellipticity (mDeg)
-25
-20
-15
-10
-5
0
5
30.00
40.00
50.00
60.00
70.00
80.00
Temperature (°C)
Cel7A
N301K
334
340
350
363
Thermal Melts of H. jecorina Cel7A and Variants

Figure 9B

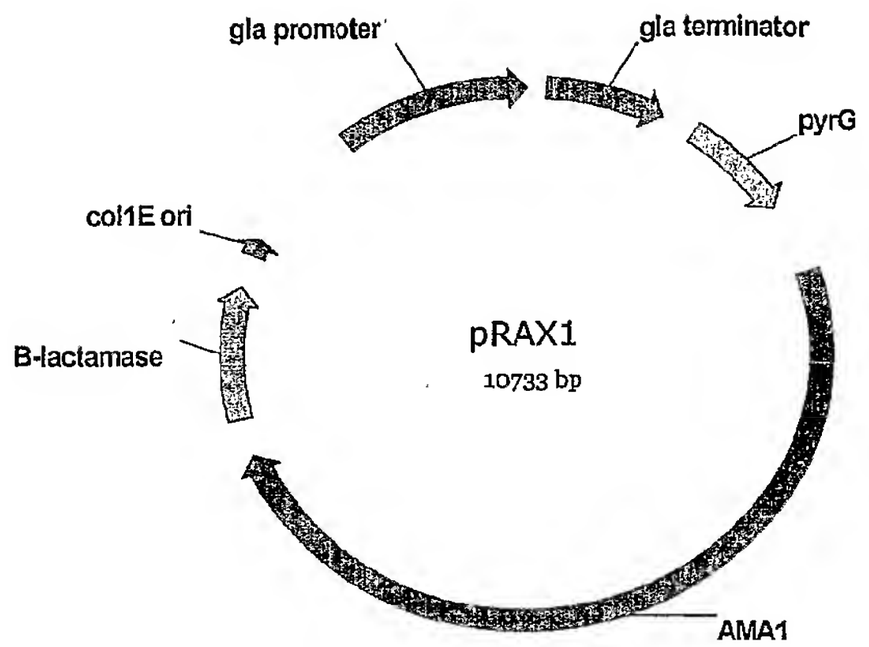


Figure 10: pRAX1

Figure 11: Destination vector pRAXdes2 for expression in *A. niger*

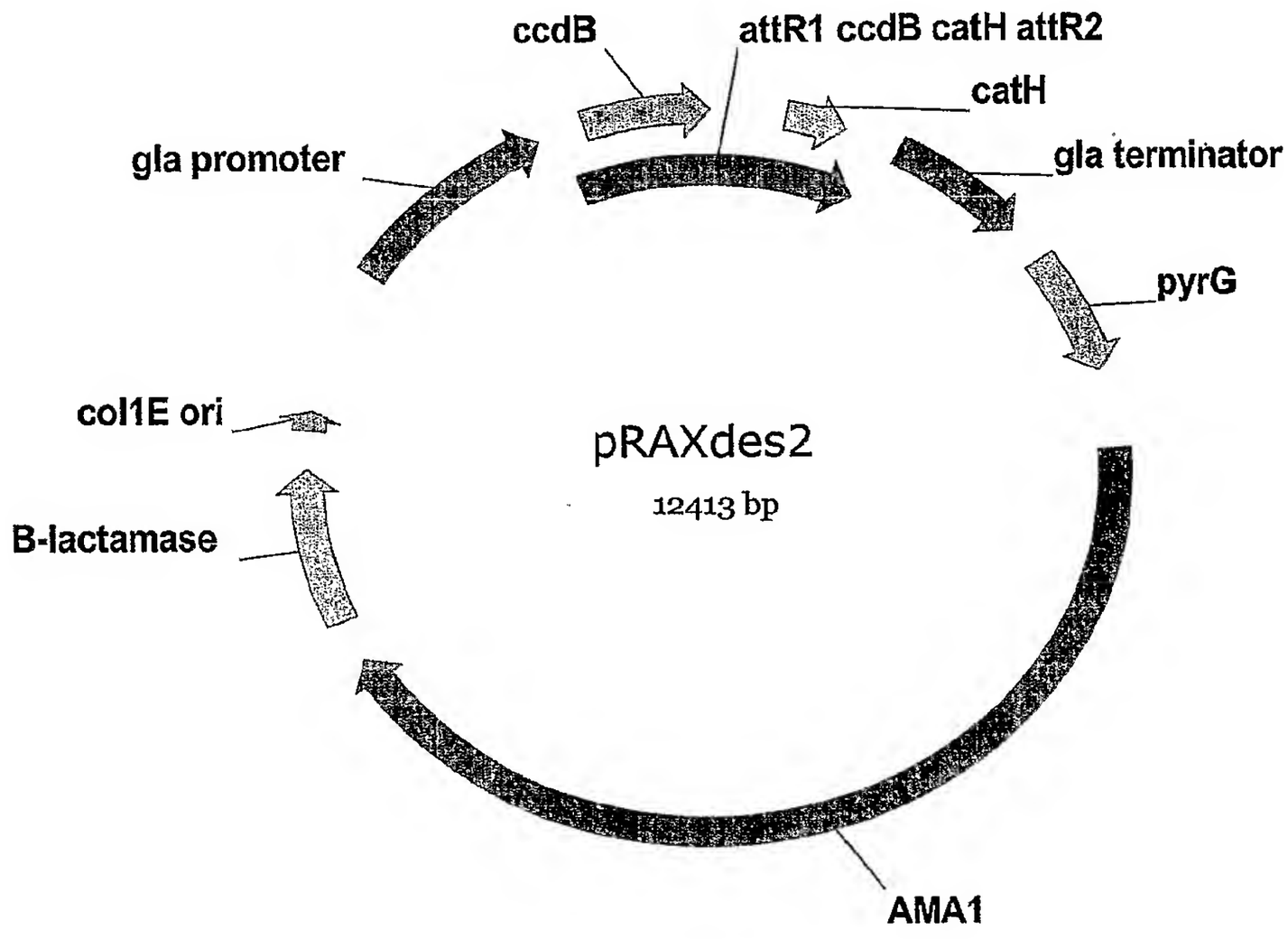


Figure 12: Replicative expression pRAXdesCBH1 vector of CBH1 genes under the control of the glucoamylase promotor.